Evolutionary clues to the molecular function of Fanconi anemia genes

Blom, E.

2006

Evolutionary clues to the molecular function of Fanconi anemia genes

Eric Blom

The research described in this thesis was performed at the Department of Clinical Genetics and Human Genetics of the VU University Medical Center, Amsterdam, the Netherlands.

Cover: Kees van der Werf

Thesis Vrije Universiteit Amsterdam
ISBN: 90-8559-158-9

Printed by: Optima Grafische Communicatie

VRIJE UNIVERSITEIT

Evolutionary clues to the molecular function of Fanconi anemia genes

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op vrijdag 21 april 2006 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Eric Blom

geboren te Woerden

promotor:	prof.dr. H. Joenje
copromotoren:	dr. F. Arwert dr. J.P. de Winter

voor mijn ouders

Leden van de leescommissie:

Dr. J. Surrallés
Universitat Autònoma de Barcelona
Spain

Prof.dr. R.F. Hoekstra
Wageningen University
the Netherlands

Prof.dr. H. te Riele
Dutch Cancer Institute, Amsterdam
the Netherlands

Prof.dr. R.H. Brakenhoff
VU University Medical Center, Amsterdam
the Netherlands

Dr. P.J.F. Snijders
VU University Medical Center, Amsterdam
the Netherlands

Dr. Q. Waisfisz
VU University Medical Center, Amsterdam
the Netherlands

Prof.dr. P. Heutink
VU University Medical Center, Amsterdam
the Netherlands

Table of contents

CHAPTER

1

General introduction and outline of thesis

Fanconi anemia

Fanconi anemia (FA) is a autosomal recessive and X-linked inherited syndrome, named after the Swiss pediatrician Guido Fanconi. In 1927, Fanconi reported the first cases of the disease and he described three brothers who had a peculiar combination of progressive aplastic anemia with several developmental anomalies [1]. After then, many other cases of this disease have been studied and the syndrome is currently characterized by progressive pancytopenia, congenital anomalies such as abnormal thumbs and radii, cancer predisposition (predominantly acute myeloid leukemia and squamous cell carcinoma), growth retardation and hyperpigmentation of the skin [2–4]. Especially due to the life-threatening bone marrow failure with a mean onset of about 8 years of age and the elevated risk for developing acute myeloid leukemia (AML), life expectancy of FA patients is reduced to an average of 25 years. The syndrome is very rare with a birth prevalence of only 5-25 cases per million newborns.

Diagnosis, cellular phenotype and differential diagnosis

Fanconi anemia patients show a considerable phenotypic variability. Some patients are not severely affected with mild anemia, no congenital anomalies and may have a life expectancy of up to 50 years, whereas others have severe congenital abnormalities and die from anemia or AML in the first decade of life. Even affected siblings or monozygotic twins may show heterogeneous congenital anomalies. Therefore, the clinical presentation may not be used for a reliable diagnosis and to discriminate FA from clinically overlapping syndromes such as VATER/VACTERL association [5], thrombocytopenia-absent radii (TAR) syndrome [6, 7] and Diamond-Blackfan anemia [8, 9]. At the cellular level, however, FA patients show a more uniform phenotype. FA patient-derived cells show spontaneous chromosome instability and in addition are hypersensitive to the chromosome-breaking and antiproliferative effects of DNA cross-linking agents such as mitomycin C (MMC) and diepoxybutane (DEB) [2, 10]. This hypersensitivity to MMC or DEB is used to diagnose FA and to discriminate the disease from the three clinically overlapping conditions that are not hypersensitive to MMC or DEB.

Although the MMC/DEB test has been proven to be a reliable tool for the diagnosis of FA, patients with other syndromes may score positive to this test as well. Nijmegen breakage syndrome (NBS) is a chromosomal instability syndrome characterized by immune deficiency, microcephaly and hypersensitivity to ionizing radiation. Since it has been reported that NBS can clinically mimic FA [11], mutation screening in the gene that is responsible for Nijmegen breakage syndrome, *NBS1*, or

a Western blot analysis of the NBS1 protein may be required to distinguish FA from NBS.

Complementation groups and genetic heterogeneity

With the availability of the MMC test to discriminate FA cells from wild type cells, it became possible to study potential locus heterogeneity of Fanconi anemia. The principle of genetic heterogeneity lies in the fact that some gene products do not perform a certain function 'on their own' but require the help of additional gene products. These gene products then cooperate in a 'biochemical pathway'. The absence of one of the components – no matter which component – then completely disrupts that pathway.

When cells from different FA complementation groups are fused, this is expected to result in a full restoration of the biochemical pathway. Although both cells miss a critical component, that component is different for each cell. Fused together, a complete set of functional genes is then present reestablishing the biochemical pathway; the hybrid cells become resistant to MMC, similar as wild type cells. However, when cells of identical complementation groups are fused, the pathway is not restored and the cells remain hypersensitive to MMC.

Extensive cell fusion studies have revealed that at least 11 complementation groups are present: FA-A, B, C, D1, D2, E, F, G, I, J, and L [12–15]. This indicates that Fanconi anemia is genetically a highly heterogeneous disorder and suggesting that at least 11 gene products may be involved in the FA biochemical pathway. However, as illustrated by the Ataxia telangiectasia (AT) disorder the rule 'one complementation group, one gene' does not always apply. In AT patient-derived cells evidence for 4 different complementation groups was found [16, 17], but mutations in a single gene, *ATM*, were found in all complementation groups suggesting that intragenic complementation must have contributed to this phenomenon.

Cloning of FA genes

In 1992, a technique was developed to clone cDNAs for Fanconi anemia by the use of functional complementation [18]. Similar to the assay used in complementation, this technique exploits the hypersensitivity to MMC that characterizes FA cells. In brief, cells belonging to a known complementation group are transfected with an episomal cDNA expression library. After selection for stable transformants, cells are treated with MMC essentially eliminating all cells that have not taken up the proper cDNA that functionally restores the FA biochemical pathway. After expansion of

these corrected cells, the episomally replicated plasmid can be extracted and analyzed by DNA sequencing.

This complementation cloning technique has proven to be highly successful and resulted in the identification of *FANCA, -C, -E, -F,* and *-G* genes that correspond to the complementation groups A, C, E, F, and G respectively [18–22]. In addition to the complementation cloning technique, conventional positional cloning resulted in identification of *FANCA* (independent of the complementation cloning of this gene) and *FANCD2* [23, 24]. Moreover, a candidate gene approach resulted in the discovery of *FANCD1* (which is identical to the breast cancer susceptibility gene *BRCA2*). The identification of *FANCL*, and very recently the identification of *FANCB* were based on protein association studies [25–27]. At present, the putative *FANCI* and *FANCJ* genes remain to be found. Very recently and after this thesis was completed, the novel complementation group FA-M, its corresponding *FANCM* gene, as well as the *FANCJ* gene were identified [28–30].

The FA biochemical pathway

With the identification of most of the FANC genes it became possible to study the FANC gene products and their role in the FA biochemical pathway. Compelling evidence has been accumulated that most of the FANC proteins, FANCA, -B, -C, -E, -F, -G, and -L, assemble in a nuclear multisubunit protein complex, termed the FA core complex. Direct interactions have been shown between FANCA and FANCG [31–33], between FANCF and FANCG [34–36], and between FANCC and FANCE [34, 37]. Although some cytoplasmic subcomplexes may be present, the largest FA core complex resides in the nucleus and has a size of 1–2 MDa [26, 27, 38–42]. Furthermore, several currently unknown components of the complexes remain to be identified, as shown in experiments by Meetei *et al.* [43].

The presence of a functional FA core complex is essential to activate the downstream FANCD2 protein by monoubiquitylation of Lys^{561} [44]. In contrast to polyubiquitylation, which targets a protein for destruction by the proteasome, monoubiquitin is a regulator of the location and activity of a diverse set of proteins [45]. In all FA patient-derived cells, except those from FA-D1 and FA-J complementation groups, this monoubiquitylation of FANCD2 is absent. Functional correction of these cell lines by transfection with the corresponding cDNAs restores the monoubiquitylation and the sensitivity to DNA cross-linking agents returns to a wild type level. Since it has been shown that FANCE directly interacts with FANCD2 [35, 37], this protein may physically link the FA core complex to FANCD2 to enable its monoubiquitylation. Except for FANCL, none of the core complex subunits harbors a catalytic domain and the RING domain of FANCL is expected to be

involved in this modification of FANCD2 [26, chapter 5 of this thesis]. It has now widely been recognized that monoubiquitylation of FANCD2 is an essential step in the FA pathway. In addition, this modification is required for further downstream reactions, such as the colocalization with BRCA1 in DNA damage induced repair foci [44]. Moreover, FANCD1/BRCA2 as well as FANCJ are currently hypothesized to function downstream of FANCD2. The precise role of FANCD1/BRCA2 within the pathway remains unclear, but it has been shown that this protein directly interacts with FANCD2 [46, 47].

Function of the FA pathway

Repair of DNA interstrand cross-links – The sensitivity of FA cells to a wide range of agents that induce DNA interstrand cross-links (ICLs) indicates that the FA pathway is involved in the repair of this type of DNA damage. Interestingly, ICLs are among the most toxic types of DNA damage and ICL repair is complicated because both strands are affected and no complementary strand is available as template for repair. It has been indicated that both nucleotide excision repair (NER) and recombinational repair are required for ICL removal and that double-strand breaks (DSBs) are produced during the process in eukaryotic cells [48]. In *Saccharomyces cerevisiae* it has been shown that during psoralen induced ICL repair NER excises the cross-linked DNA on both strands, producing a DSB intermediate, which is then repaired through homologous recombination (HR) [49, 50].

Emerging evidence suggests that the FA pathway of ICL repair may act mainly during the S-phase of the cell cycle and after DSBs have been formed. The activation of FANCD2 during S-phase and the S-phase specific colocalization of FANCD2 with BRCA1 and RAD51 [51] suggests that ICL repair mediated by the FA pathway may be specific to this phase of the cell cycle. In has been shown that psoralen-treated primary fibroblasts arrest in the late S-phase regardless of when the damage was introduced during the cell cycle [52]. During DNA synthesis, replication forks that encounter an ICL may collapse and induce a cell cycle checkpoint in order to allow sufficient time to repair the ICL. Moreover, replicative stress-inducing agents as aphidicolin and hydroxyurea also induce strong monoubiquitylation of FANCD2 [53].

Hypersensitivity of FA cells to ICLs is unlikely due to a defect in the initial step of ICL repair. Pulsed-field gel electrophoresis and single-cell electrophoresis of MMC-treated cells show that FA cells form DSBs and unhook MMC-induced ICL similarly as wild type cells [54]. This indicates that the FA pathway may function after the generation of DSBs has occurred [55]. The colocalization of activated FANCD2 with the homologous recombination repair proteins BRCA1 and RAD51 [51] as well as

FANCD1 being identical to BRCA2 [25] suggests that these generated DSBs may be repaired through homologous recombination. However, as recently suggested by Nakanishi *et al.* [56] the FA pathway may – besides its role in promoting homology-directed repair of DSBs – also promote single-strand annealing which is a alternative route for repair of DSBs.

Other functions of the FA pathway – Previous research has indicated that the FANC proteins may have additional functions. Primary FA cells show a marked sensitivity to a non-physiologically high level of oxygen resulting in chromosomal breakage [57] and arrest in the G2 phase of the cell cycle [58], in contrast to wild type cells. Another study shows that *Fancc*$^{-/-}$ *Sod1*$^{-/-}$ double knockout mice have hematopoietic deficiency but not *Fancc*$^{-/-}$ and *Sod1*$^{-/-}$ single knockouts [59]. While this does not necessarily indicate a direct role of the FA pathway in oxygen metabolism, it may indicate that metabolically generated oxygen radicals may be more toxic to FA cells than to wild type cells.

The FANCC protein has been shown to interact with NADPH cytochrome P450 reductase, the transcription factor STAT1, and glutathione S-transferase P1-1 [60–62], suggesting multiple functions for this protein. In addition, extensive yeast two-hybrid (Y2H) screening has identified 69 different interactors of FANCA, FANCC, and FANCG, including DAXX, Ran, IκBγ, USP14, SNX5, and FAZF [63]. Most of these interacting proteins are associated with transcription regulation, signaling, oxidative metabolism, and intracellular transport. However, while the FA core complex purification method by Meetei *et al.* [43] has proven its applicability resulting in the identification of three novel FANC genes (*FANCB, FANCL*, and *FANCM*) the reliability of the Y2H dataset is still remaining obscure and may contain false positives.

Animal models for FA

Since the *FANC* genes are conserved throughout vertebrate evolution, several animal models have been used to study Fanconi anemia. *Fanca*$^{-/-}$, *Fancc*$^{-/-}$, *Fancg*$^{-/-}$, *Fancl/Pog*$^{-/-}$, and *Fanca*$^{-/-}$*/Fancc*$^{-/-}$ mutant mice show no developmental abnormalities, bone marrow failure or cancer proneness but do show a decreased fertility, spontaneous chromosomal instability at the cellular level and hypersensitivity to DNA cross linking agents [64–70]. In contrast, Fanconi anemia complementation group D2 knockout mice have been reported to show a more severe phenotype including microphthalmia, perinatal lethality, and epithelial cancers [71]. The chicken B cell line DT40 has also been used for studying the FA biochemical pathway and *Fancg*$^{-/-}$, *Fancc*$^{-/-}$, and *Fancd2*$^{-/-}$ knockout DT40 cells have

been reported [72–74]. These chicken cells show increased sensitivity to DNA cross linking agents and exhibit increased MMC and radiation-induced chromosomal breakage. Finally, Liu *et al.* [75] showed that downregulation of Fancd2 during zebrafish embryogenesis caused defects similar to those found in FA such as growth retardation, microcephaly, and microopthalmia.

Aim and outline of this thesis

Although the majority of the FANC genes have been identified to date, still little is known about the biochemistry of the FA pathway and how it is involved in ICL repair. While most other protein sequences contain conserved domains providing clues for their biochemical function, the FANC proteins are all 'orphans' showing no homology to other proteins making a prediction about their function difficult. In addition, the very limited conserved FANC proteins have no homologs in important model species such as *Escherichia coli* and yeast in which ICL repair pathways have been studied in great detail.

The aim of this thesis was to study the FANC proteins in evolutionary distantly related species, e.g. chicken, clawed frog, fish, and fruit fly. Amino acids that are conserved between these species may be functionally important and could potentially provide clues about the molecular action of the FANC proteins.

In **chapter 2** the principle of orphan gene products, protein domains, and evolutionary conservation is further explained. In this chapter, it is discussed how a bioinformatic approach could provide important information about the biochemical function of a protein. In a preliminary analysis, evidence was found that FANCG contains at least 2 tetratricopeptide repeat (TPR) motifs. The FANCG sequence was further studied in **chapter 3** and a total of 7 TPRs were identified. The TPR motif is a protein-protein interaction domain and TPR-containing proteins often function as scaffolds mediating the assembly of multisubunit protein complexes. The importance of the TPRs in FANCG was studied by mutational analysis which confirmed that the TPRs function as protein-protein interaction modules and are required for a functional FA pathway. In **chapter 4**, the FANCF sequence was studied but no protein domain was identified. Extensive mutational analysis suggests that this protein functions as an important adaptor molecule within the FA core complex. All other FANC protein sequences were also examined (**chapter 5**) and some evidence was obtained that FANCA and FANCC may contain HEAT repeats. Similarly as for the FANCG protein, FANCA and FANCC may function as scaffolds in the assembly of the FA core complex. Furthermore, FANCD2 may be a myosin heavy chain-like molecule. Careful analysis of FANCL revealed that it does

not contain a PHD domain – as previously suggested by others [26] – but a RING domain required for its ubiquitin ligase activity.

Since in non-vertebrates such as the fruit fly *Drosophila melanogaster* the FA core complex subunits are absent, it was not possible to study their sequences in these species. However, the recent identification of *FANCL*, which – in contrast to the other FANC core complex subunits – is also present in the fruit fly genome, allowed dissection of the putative FA pathway in this species and in **chapter 6** the first preliminary data are presented.

References

1 Fanconi G (1927). Familiäre infantile perniziosaartige Anämia (perniziöses Blutbild und Konstitution). *Jahrb Kinderh* **117:** 257–280.

2 Joenje H, Patel KJ (2001). The emerging genetic and molecular basis of Fanconi anaemia. *Nat Rev Genet* **2:** 446–457.

3 Tischkowitz MD, Hodgson SV (2003). Fanconi anaemia. *J Med Genet* **40:** 1–10.

4 D'Andrea AD, Grompe M (2003). The Fanconi anaemia/BRCA pathway. *Nat Rev Cancer* **3:** 23–34.

5 Corsello G (1992) VATER/VACTERL association: clinical variability and expanding phenotype including laryngeal stenosis. *Am J Med Genet* **44**: 813–815.

6 Hall JG (1987). Thrombocytopenia and absent radius (TAR) syndrome. *J Med Genet* **24:** 79–83.

7 Hedberg VA, Lipton JM (1988). Thrombocytopenia with absent radii. A review of 100 cases. *Am J Pediatr Hematol Oncol* **10:** 51–64.

8 D'Andrea AD, Dahl N, Guinan EC, Shimamura A (2002). Marrow failure. *Hematology (Am Soc Hematol Educ Program)* 58–72.

9 Dianzani I, Garelli E, Ramenghi U (2000). Diamond-Blackfan Anaemia: an overview. *Paediatr Drugs* **2:** 345–355.

10 Sasaki MS, Tonomura A (1973). A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross-linking agents. *Cancer Res* **33:** 1829–1836.

11 New HV *et al.* (2005). Nijmegen breakage syndrome diagnosed as Fanconi anaemia. *Pediatr Blood Cancer* **44:** 494–499.

12 Strathdee CA, Duncan AM, Buchwald M (1992). Evidence for at least four Fanconi anaemia genes including *FACC* on chromosome 9. *Nat Genet* **1:** 196–198.

13 Joenje H *et al.* (1997). Evidence for at least eight Fanconi anemia genes. *Am. J. Hum Genet* **61:** 940–944.

14 Joenje H *et al.* (2000). Complementation analysis in Fanconi anemia: assignment of the reference FA-H patient to group A. *Am J Hum Genet* **67:** 759–762.

15 Levitus M *et al.* (2004). Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes. *Blood* **103:** 2498–2503.

16 Jaspers NG, Gatti RA, Baan C, Linssen PC, Bootsma D (1988). Genetic complementation analysis of ataxia telangiectasia and Nijmegen breakage syndrome: a survey of 50 patients. *Cytogenet Cell Genet* **49:** 259–263.

17 Savitsky K *et al.* (1995). A single ataxia telangiectasia gene with a product similar to PI-3 kinase. *Science* **268:** 1749–1753.

18 Strathdee CA, Gavish H, Shannon WR, Buchwald M (1992). Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **356:** 763–767.

19 Lo Ten Foe JR *et al.* (1996). Expression cloning of a cDNA for the major Fanconi anaemia gene, *FAA*. *Nat Genet* **14:** 320–323.
20 de Winter JP *et al.* (1998). The Fanconi anaemia group G gene *FANCG* is identical with *XRCC9*. *Nat Genet* **20:** 281–283.
21 de Winter JP *et al.* (2000). The Fanconi anaemia gene *FANCF* encodes a novel protein with homology to ROM. *Nat Genet* **24:** 15–16.
22 de Winter JP *et al.* (2000). Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am J Hum Genet* **67:** 1306–1308.
23 The Fanconi anaemia/breast cancer consortium (1996). Positional cloning of the Fanconi anaemia group A gene. *Nat Genet* **14:** 324–328.
24 Timmers C *et al.* (2001). Positional cloning of a novel Fanconi anemia gene, *FANCD2*. *Mol Cell.* **7:** 241–248.
25 Howlett NG *et al.* (2002). Biallelic inactivation of *BRCA2* in Fanconi anemia. *Science* **297:** 606–609.
26 Meetei AR *et al.* (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35:** 165–170.
27 Meetei AR *et al.* (2004). X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet* **36:** 1219–1224.
28 Levran O *et al.* (2005). The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia. *Nat Genet* **37:** 931–933.
29 Levitus M *et al.* (2005). The DNA helicase BRIP1 is defective in Fanconi anemia complementation group J. *Nat Genet* **37:** 934–935.
30 Meetei AR *et al.* (2005). A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet* **37:** 958–963.
31 Waisfisz Q *et al.* (1999). A physical complex of the Fanconi anemia proteins FANCG/XRCC9 and FANCA. *Proc Natl Acad Sci USA* **96:** 10320–10325.
32 Kruyt FA, Abou-Zahr F, Mok H, Youssoufian H (1999). Resistance to mitomycin C requires direct interaction between the Fanconi anemia proteins FANCA and FANCG in the nucleus through an arginine-rich domain. *J Biol Chem* **274:** 34212–34218.
33 Huber PA, Medhurst AL, Youssoufian H, Mathew CG (2000). Investigation of Fanconi anemia protein interactions by yeast two-hybrid analysis. *Biochem Biophys Res Commun* **268:** 73–77.
34 Medhurst AL, Huber PA, Waisfisz Q, de Winter JP, Mathew CG (2001). Direct interactions of the five known Fanconi anaemia proteins suggest a common functional pathway. *Hum Mol Genet* **10:** 423–429.
35 Gordon SM, Buchwald M (2003). Fanconi anemia protein complex: mapping protein interactions in the yeast 2- and 3-hybrid systems. *Blood* **102:** 136–141.

36 Léveillé F *et al.* (2004). The Fanconi anemia gene product FANCF is a flexible adaptor protein. *J Biol Chem* **279:** 39421–39430.
37 Pace P *et al.* (2002). FANCE: the link between Fanconi anaemia complex assembly and activity. *EMBO J* **21:** 3414–3423.
38 Kupfer GM, Naf D, Suliman A, Pulsipher M, D'Andrea AD (1997). The Fanconi anaemia proteins, FAA and FAC, interact to form a nuclear complex. *Nat Genet* **17:** 487–490.
39 Garcia-Higuera I, Kuang Y, Naf D, Wasik J, D'Andrea AD (1999). Fanconi anemia proteins FANCA, FANCC, and FANCG/XRCC9 interact in a functional nuclear complex. *Mol Cell Biol* **19:** 4866–4873.
40 de Winter JP *et al.* (2000). The Fanconi anemia protein FANCF forms a nuclear complex with FANCA, FANCC and FANCG. *Hum Mol Genet* **9:** 2665–2674.
41 Taniguchi T, D'Andrea AD (2002). The Fanconi anemia protein, FANCE, promotes the nuclear accumulation of FANCC. *Blood* **100:** 2457–2462.
42 Thomashevski A *et al.* (2004). The Fanconi anemia core complex forms four complexes of different sizes in different subcellular compartments. *J Biol Chem* **279:** 26201–26209.
43 Meetei AR *et al.* (2003). A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome. *Mol Cell Biol* **23:** 3417–3426.
44 Garcia-Higuera I *et al.* (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7:** 249–262.
45 Hicke L (2001). Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol* **3:** 195–201.
46 Hussain S *et al.* (2004). Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum Mol Genet* **13:** 1241–1248.
47 Wang X, Andreassen PR, D'Andrea AD (2004). Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol Cell Biol* **24:** 5850–5862.
48 Dronkert ML, Kanaar R (2001). Repair of DNA interstrand cross-links. *Mutat Res* **486:** 217–247.
49 Jachymczyk WJ, von Borstel RC, Mowat MR, Hastings PJ (1981). Repair of interstrand cross-links in DNA of Saccharomyces cerevisiae requires two systems for DNA repair: the RAD3 system and the RAD51 system. *Mol Gen Genet* **182:** 196–205.
50 Miller RD, Prakash L, Prakash S (1982). Genetic control of excision of Saccharomyces cerevisiae interstrand DNA cross-links induced by psoralen plus near-UV light. *Mol Cell Biol* **2:** 939–948.

51 Taniguchi T *et al.* (2002). S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood* **100:** 2414–2420.
52 Akkari YM, Bateman RL, Reifsteck CA, Olson SB, Grompe M (2000). DNA replication is required To elicit cellular responses to psoralen-induced DNA interstrand cross-links. *Mol Cell Biol* **20:** 8283–8289.
53 Howlett NG, Taniguchi T, Durkin SG, D'Andrea AD, Glover TW (2005). The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Hum Mol Genet* **14:** 693–701.
54 Pichierri P, Averbeck D, Rosselli F (2002). DNA cross-link-dependent RAD50/MRE11/NBS1 subnuclear assembly requires the Fanconi anemia C protein. *Hum Mol Genet* **11:** 2531–2546.
55 Rothfuss A, Grompe M (2004). Repair kinetics of genomic interstrand DNA cross-links: evidence for DNA double-strand break-dependent activation of the Fanconi anemia/BRCA pathway. *Mol Cell Biol* **24:** 123–134.
56 Nakanishi K *et al.* (2005). Human Fanconi anemia monoubiquitination pathway promotes homologous DNA repair. *Proc Natl Acad Sci USA* **102:** 1110–1115.
57 Joenje H, Arwert F, Eriksson AW, de Koning H, Oostra AB (1981). Oxygen-dependence of chromosomal aberrations in Fanconi's anaemia. *Nature* **290:** 142–143.
58 Kubbies M, Schindler D, Hoehn H, Schinzel A, Rabinovitch PS (1985). Endogenous blockage and delay of the chromosome cycle despite normal recruitment and growth phase explain poor proliferation and frequent edomitosis in Fanconi anemia cells. *Am J Hum Genet* **37:** 1022–1030.
59 Hadjur S *et al.* (2001). Defective hematopoiesis and hepatic steatosis in mice with combined deficiencies of the genes encoding Fancc and Cu/Zn superoxide dismutase. *Blood* **98:** 1003–1011.
60 Kruyt FAE *et al.* (1998). Abnormal microsomal detoxification implicated in Fanconi anemia group C by interaction of the FAC protein with NADPH cytochrome P450 reductase. *Blood* **92:** 3050–3056.
61 Pang Q *et al.* (2000). The Fanconi anemia protein FANCC binds to and facilitates the activation of STAT1 by gamma interferon and hematopoietic growth factors. *Mol Cell Biol* **20:** 4724–4735.
62 Cumming RC *et al.* (2001). Fanconi anemia group C protein prevents apoptosis in hematopoietic cells through redox regulation of GSTP1. *Nat Med* **7:** 814–820.
63 Reuter TY *et al.* (2003). Yeast two-hybrid screens imply involvement of Fanconi anemia proteins in transcription regulation, cell signaling, oxidative metabolism, and cellular transport. *Exp Cell Res* **289:** 211–221.

64 Chen M *et al.* (1996). Inactivation of Fac in mice produces inducible chromosomal instability and reduced fertility reminiscent of Fanconi anaemia. *Nat Genet* **12:** 448–451.
65 Whitney MA *et al.* (1996). Germ cell defects and hematopoietic hypersensitivity to gamma-interferon in mice with a targeted disruption of the Fanconi anemia C gene. *Blood* **88:** 49–58.
66 Chen NC *et al.* (2000). Mice with a targeted disruption of the Fanconi anemia homolog Fanca. *Hum Mol Genet* **9:** 1805–1811.
67 Yang Y *et al.* (2001). Targeted disruption of the murine Fanconi anemia gene, *Fancg/Xrcc9*. *Blood* **98:** 3435–3440.
68 Koomen M *et al.* (2002). Reduced fertility and hypersensitivity to mitomycin C characterize *Fancg/Xrcc9* null mice. *Hum Mol Genet* **11:** 273–281.
69 Agoulnik AI *et al.* (2002). A novel gene, Pog, is necessary for primordial germ cell proliferation in the mouse and underlies the germ cell deficient mutation, gcd. *Hum Mol Genet* **11:** 3047–3053.
70 Wong JCY *et al.* (2003). Targeted disruption of exons 1 to 6 of the Fanconi anemia A gene leads to growth retardation, strain-specific microphthalmia, meiotic defects and primordial germ cell hypoplasia. *Hum Mol Genet* **12:** 2063–2076.
71 Houghtaling S *et al.* Epithelial cancer in Fanconia anemia complementation group D2 (*Fancd2*) knockout mice. *Genes Dev* **17:** 2021–2035.
72 Yamamoto K *et al.* (2003). Fanconi anemia FANCG protein in mitigating radiation- and enzyme-induced DNA double-strand breaks by homologous recombination in vertebrate cells. *Mol Cell Biol* **23:** 5421–5430.
73 Niedzwiedz W *et al.* (2004). The Fanconi anaemia gene *FANCC* promotes homologous recombination and error-prone DNA repair. *Mol Cell* **15:** 607–620.
74 Yamamoto K *et al.* (2005). Fanconi anemia protein FANCD2 promotes immunoglobulin gene conversion and DNA repair through a mechanism related to homologous recombination. *Mol Cell Biol* **25:** 34–43.
75 Liu TX *et al.* (2003). Knockdown of zebrafish Fancd2 causes developmental abnormalities via p53-dependent apoptosis. *Dev Cell* **5:** 903–914.

CHAPTER

2

Evolutionary clues to the molecular function of Fanconi anemia genes

Fanconi anemia (FA) is an autosomal recessively inherited disease with diverse clinical symptoms including developmental anomalies, predisposition to neoplasia, and a deficiency of hematopoietic stem cells resulting in progressive aplastic anemia. FA is genetically heterogeneous with at least 8 genes being implicated on the basis of functional complementation studies. To date, six FA genes are known: *FANCA, FANCC, FANCD2, FANCE, FANCF* and *FANCG*, all of which encode orphan proteins sharing no homology to each other nor to any other known protein. In addition, they do not appear to possess any domains with homology to currently known protein domains, which makes a prediction about their molecular action difficult. Studying the molecular evolution of FA genes and their products using sensitive database search methods such as PSI-BLAST may provide novel insight into the nature of the FA pathway and its relationship to hematopoiesis, embryonic development and the origin of malignancies. Preliminary results of such an approach show that at least one FA protein, FANCG, may contain a known domain, suggesting that this protein is a member of the family of tetratricopeptide repeat-containing proteins.

Eric Blom, Henri J. van de Vrugt, Johan P. de Winter,
Fré Arwert & Hans Joenje

Acta Haematol 2002; 108: 231–236

Introduction

Fanconi anemia (FA) is a rare autosomal recessive disorder characterized by multiple congenital anomalies, progressive bone marrow failure, and cancer predisposition, especially acute myeloid leukemia (AML) and squamous cell carcinomas [1, 2]. Cells derived from FA patients show spontaneous chromosomal instability and a characteristic hypersensitivity to DNA cross-linking agents, such as mitomycin C and diepoxybutane [3, 4]. With at least eight distinct disease genes predicted from somatic cell hybridization studies, FA is genetically highly heterogeneous [5–7]. To date, six FA genes have been cloned [8–14], but little is known about the molecular action of their gene products (table 1). All FA genes encode orphan proteins that share no homology to each other nor to any other known protein. In addition, no functional domains have so far been detected in these proteins, making a prediction about their molecular function difficult. In normal cells, the FA proteins FANCA, FANCC, FANCG, and FANCF assemble in a multi-subunit nuclear core complex [15–18], which suggests that protein-protein binding domains may be detectable within these proteins. The FA protein core complex, however, fails to be formed in cell lines derived from complementation groups B and E, suggesting an upstream action for the FANCB and FANCE proteins in the FA pathway [19].

Recently, the *FANCD2* gene was cloned [14] and appears to encode yet another orphan protein. Unlike the other FA proteins, FANCD2 is conserved in nonvertebrate animals, such as *Drosophila melanogaster* and even in the plant *Arabidopsis thaliana*, indicating that the FA pathway may be conserved throughout eukaryote evolution. In the presence of a functional FA protein core complex, the FANCD2 protein is activated to a monoubiquitinated isoform that accumulates in nuclear foci in response to DNA damaging agents and colocalizes with the breast cancer susceptibility protein BRCA1 [20].

Since during evolution new proteins typically arise via gene duplication or exon shuffling [21], it seems very unlikely that FA proteins are without any similarity to other proteins. In gene duplication, the complete gene is duplicated, after which one copy retains its original function whereas the other copy is available for acquiring a new function by the process of random mutagenesis or exon shuffling in combination with selection. Alternatively, the duplicated gene may become inactivated, e.g. by a stop codon, and as a result this gene may remain in the genome as a *pseudogene*. For example, myoglobin as well as the α- and β-families of hemoglobin have all arisen from one ancestral gene, while also several pseudogenes are present in the human genome. Evolution via exon shuffling occurs when a new exon is inserted into an existing gene or when an exon is

duplicated within the same gene. Since exons often represent functional protein domains (regions that perform a specific function), acquisition of a new exon could enhance protein function. Tissue-type specific plasminogen activator is an example of a protein that is composed of different functional domains. It contains a finger domain, a growth factor domain, two kringle domains and an enzymatic serine protease domain and is therefore a good example of a mosaic protein, which has presumably arisen by exon shuffling.

Finding similarities between proteins can be an important starting point for the prediction of molecular function, which can then be experimentally tested. However, the recently identified FA proteins: FANCA, -C, -D2, -E, -F, and -G are all unique proteins that lack statistically significant similarities to any of currently known protein domains making predictions about their molecular function difficult. One strategy to bypass this obstacle could be to study the evolutionary *conservation* among different species and subsequently applying sensitive database search methods in order to identify functionally important domains. Such a strategy would require first that orthologs of FA genes in nonmammalian species, such as fish and *Xenopus,* are determined. Orthologs are genes that originated from the same ancestral gene at an earlier evolutionary time and are likely to perform the same or a similar function. A multiple alignment is then prepared of these orthologous sequences after which a *specific* rather than a *standard* score model is utilized for the protein in question. For each position, each amino acid is given a score depending on how likely it is to occur. This scoring model can then be used for searching the database for similarities in a more sensitive way than the BLAST algorithm. PSI-BLAST (position-specific iterated blast) is an example of a program that is easy to use and applies a position-specific scoring matrix. It performs an iterative search in which sequences found in one searching round are used to build a score model for the next round. We applied this approach focusing on the FANCG protein, which appears to be conserved in the zebrafish. Searching the nonredundant database at NCBI with the C-terminus of zebrafish Fancg using PSI-BLAST revealed a previously unknown similarity to a putative mouse and a *Drosophila* protein. This similarity is most likely based on the presence of one or more tetratricopeptide repeats (TPRs) [22, 23] in the C-terminus of FANCG. The TPR is a degenerate 34-amino-acid repeated motif with only few loosely conserved consensus residues and is thought to mediate protein-protein interaction. The presence of one or more TPR motifs in the C-terminus of FANCG may be responsible for its interaction with other proteins in the FA protein core complex.

Materials and methods

Database searches were performed using the BLAST and PSI-BLAST algorithms [24] at the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/BLAST/) using the *protein* sequences of the cloned FA genes as queries. In general, BLAST searches via a protein sequence (i.e. protein to protein database by blastp, nucleotide to protein database by blastx, or protein to nucleotide database by tblastn) are more sensitive than a nucleotide query to nucleotide database (blastn) because they avoid the problem of (functionally silent) sequence divergence due to codon degeneracy. In addition, decreasing the word size parameter from 3 to 2 tends to enhance sensitivity. Identification of possible domains was investigated using Pfam 6.4 (May 2001, 2866 models) [25] at http://pfam.wustl.edu/. Pfam is a collection of multiple alignments and profile hidden Markov models (HMMs) of protein domain families. Protein sequences can be searched against these profile HMMs to locate regions of the sequence that belong to domain families. Multiple alignments were prepared by using the ClustalW (www.ebi.ac.uk/clustalw/) and Boxshade (www.ch.embnet.org/software/BOX_form.html) programs.

Results and discussion

Evolution of FA Proteins – Between mouse and human FA proteins, only 55–80% of amino acids have remained identical during evolution (table 1). This percentage is further decreased with greater evolutionary distance, such as for human-*Xenopus* and human-fish (Fig. 1). Searching for homologs in nonvertebrate species, such as *D. melanogaster*, has failed, however. This indicates that apparently too few amino acids have remained identical between human and *Drosophila* to result in a statistically significant BLAST hit, or that FA proteins just do not exist in lower species. An interesting exception is the FANCD2 protein, which is conserved in *Caenorhabditis elegans*, *D. melanogaster* and even plants *(A. thaliana)*, which indicates that at least one FA protein was preserved during evolution. Since FA proteins are thought to function in one pathway, why is this particular FA protein more conserved than the others? Perhaps FANCD2 plays a role in a more ancient pathway on which, later in evolution, the action of the non-D2 proteins has been superimposed. In this scenario, a new regulatory mechanism for FANCD2 activity might have developed during vertebrate evolution. Alternatively, non-D2 proteins do exist in nonvertebrates, but due to their low level of amino acid conservation are undetectable, which essentially means that their BLAST scores do not reach statistical significance.

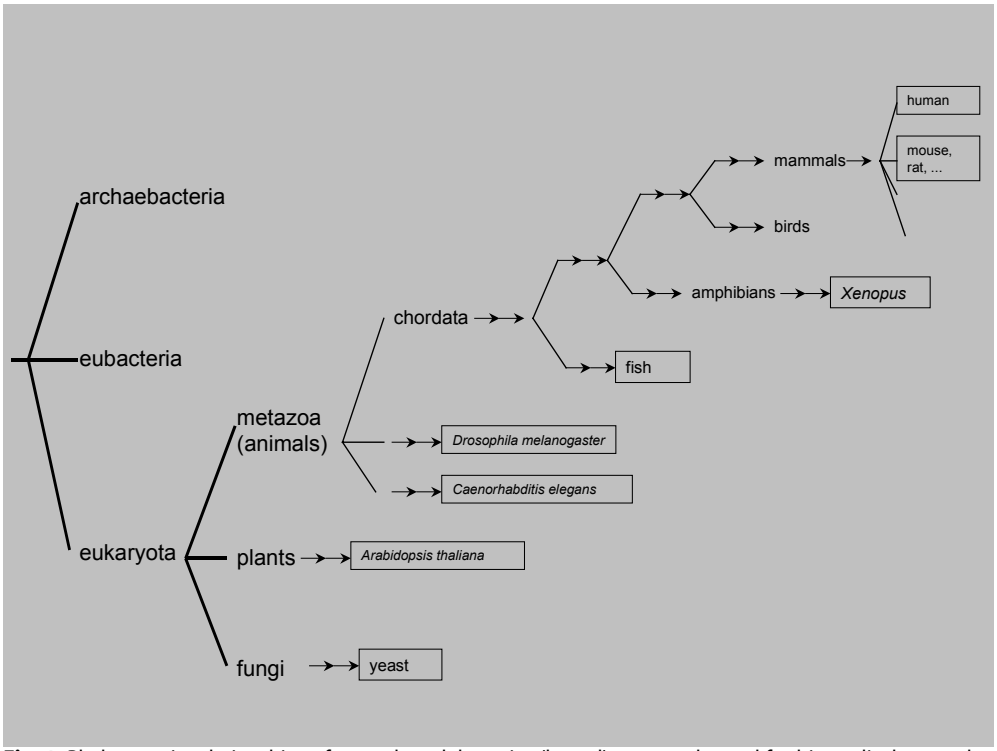


Fig. 1. Phylogenetic relationships of several model species (boxed) commonly used for biomedical research. From these species, complete genome sequences or at least EST sequences, are available that can be used for studying the evolution of the FA proteins. Evolutionary conservation of FA proteins appears to be limited to the chordata. However, the FANCD2 protein is conserved in metazoa and in the plant *A. thaliana*, indicating that at least part of the FA pathway may be conserved in all eukaryote species.

Secondary structure (distribution of α-helices and β-sheets in a protein) and tertiary structure (overall 3D structure of the protein) are generally more preserved than primary structure (amino acids), indicating that many amino acid substitutions do not change the global 3D structure of the protein and therefore do not alter protein function. This is proven by the fact that cDNAs encoding mouse *Fanca*, *Fancc*, and *Fancg* [26–28, H.J. van de Vrugt, unpubl. data], are all able to complement the corresponding human FA cells, despite an amino acid sequence identity of only 62–72% with the human sequence. Research to identify non-D2 homologues in nonvertebrates, such as *Drosophila*, should focus on database search algorithms, such as PSI-BLAST and Profile searching, which are more sensitive than the commonly used BLAST algorithm. Here, we describe preliminary results obtained by this approach applied to the FANCG protein.

Table 1. FA genes and their products

Gene	Locus	Exons	Protein kDa	Amino acids	Patients %	Evolutionary conservation	Identity human versus mouse protein sequence, %
FANCA	16q24.3	43	163	1455	~70	Fish	66
FANCB	unknown	-	-	-	rare	-	-
FANCC	9q22.3	14	63	558	~10	Fish	62[a]
FANCD1	unknown	-	-	-	rare	-	-
FANCD2	3p25.3	44	162	1451	rare	A. thaliana, C. elegans, D. melanogaster	~80[b]
FANCE	6p21.3	10	60	536	rare	Fish	65
FANCF	11p15	1	42	374	rare	Fish	~55[b]
FANCG	9p13	14	70	622	~10	Fish	72

[a] The mouse sequence contains an extra exon. For comparison: identity human-rat is 66%.
[b] This is an estimation based on EST sequences in the NCBI database.

FANCG Shows Homology to the TPR Motif – The FANCG protein appears to be conserved at least down to the ray-finned fishes. Searching the NCBI EST_others database [which includes expressed sequence tags (ESTs) derived only from non-mammalian species] using BLAST revealed a weak but significant similarity (E-value ~$1{\cdot}10^{-4}$, word size parameter set at 2 rather than 3) with a zebrafish EST clone (Acc.No. BG302995), which represents the C-terminus of FANCG (Fig. 2). In order to determine whether a BLAST hit is significant, the algorithm calculates an Expect value (E-value). The Expect value (E) is a parameter that describes the number of hits one can expect just by chance when searching a particular database, thus providing an estimate of whether a hit is significant or not. When the protein sequence of this zebrafish EST is used as a query in a PSI-BLAST search, the human FANCG is detected in the first iteration and is then used for creating a position-specific score matrix. In the second iteration, a weak but significant similarity with a putative mouse protein (Acc.No. BAB28115, E = $1{\cdot}10^{-4}$) was found and in the third iteration a putative *Drosophila* protein CG6621 (Acc.No. AAF54579, E = $6{\cdot}10^{-7}$). This indicates that at least *part* of the FANCG protein must also be present in other proteins. This similarity between these proteins is most likely based on the presence of one or more TPRs in these proteins (Fig. 3). In all these proteins, a TPR motif is predicted by Pfam. The TPR motif is present in a wide variety of proteins including e.g. the subunits cdc16, cdc23, and cdc27 of the multiprotein E3 ubiquitin ligase 'anaphase promoting' complex (APC), several transcription factors, the PKR protein kinase inhibitor, and peroxisomal as well as mitochondrial import proteins [22, 23]. It is usually present in tandem arrays of 3–16 motifs suggesting that FANCG may contain several such motifs. Sequencing the full length zebrafish *Fancg* gene may

help to pinpoint the exact locations of additional TPR motifs in human FANCG. Indeed, Pfam predicts an additional TPR motif in zebrafish Fancg C-terminus immediately downstream of the first one, but not in the human sequence. An additional TPR motif is predicted at position 246–279 in the human protein, while in both the mouse and *Drosophila* putative proteins a total of three repeats are predicted. It should be noted that both putative mouse and *Drosophila* proteins are not homologs of FANCG, but like FANCG, are predicted to contain several TPR motifs.

Tandem arrays of TPR motifs are thought to be involved in the assembly of multiprotein complexes, suggesting that TPR motifs in FANCG may play a role in the assembly of the FA protein core complex. Clinical evidence for the importance of this motif at position 514–547 in FANCG comes from the identification of two pathogenic missense mutations in this region [29]. The TPR motif is 34 amino acids long with only few loosely conserved residues (Fig. 3) that are conserved in terms of their size, hydrophobicity and spacing. The missense mutation G521E [29] alters a conserved glycine residue and presumably disrupts the TPR motif (Fig. 3). In addition, Nakanishi et al. [29] showed that this mutant protein is not detectable by immunoblotting suggesting that the protein or mRNA is unstable. Missense mutation G546R is expressed at normal levels, but fails to correct the cellular defect of FA-G cells, which confirms that this motif may be important for the function of FANCG.

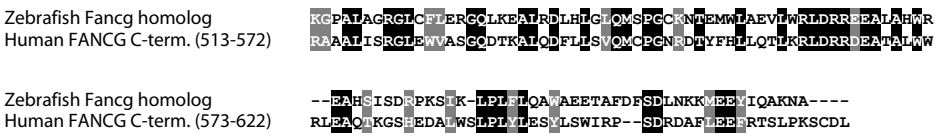


Fig. 2. Alignment of human FANCG and zebrafish Fancg homolog. Zebrafish sequence was derived from NCBI BG302995 entry (EST sequence). Since the full-length zebrafish sequence is not available yet, only the C-terminal region is shown. Alignment was prepared by using the Boxshade program.

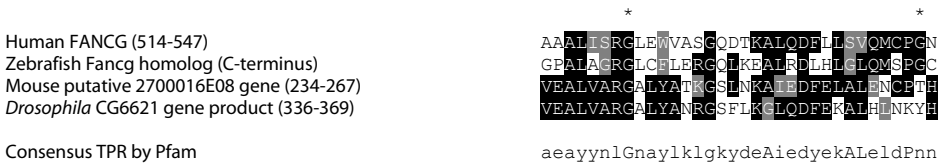


Fig. 3. Comparison of putative TPR motifs in human FANCG, zebrafish Fancg homolog (sequence derived from a EST sequence, NCBI entry BG302995), mouse putative 2700016E08 gene (Acc.no. NP_080254 or BAB28115), and *Drosophila* CG6621 gene product (Acc.no. AAF54579). In all proteins, a TPR motif is predicted by Pfam with an expected value above the trusted cut-off value. *Note that G521E is a pathogenic mutation in human FANCG which presumably destabilizes the protein or mRNA [29]. In addition, the G546R missense mutation is expressed at normal levels, but fails to complement FA-G cells [29]. Alignment was prepared by using the Boxshade program.

Conclusion

We conclude that using database search algorithms such as PSI-BLAST may be a useful approach to classify FA gene products into known protein families. Preliminary data show that at least one FA protein, FANCG, appears to contain a known domain, indicating that this protein is a member of the TPR-containing family of proteins. Since for this approach sequence information for vertebrate orthologs of the FA proteins is required, research should first focus on the cloning and sequencing of these orthologs. Fish and *Xenopus* deserve priority, because genomic or EST sequence data are already available which is likely to facilitate the cloning of the cDNAs. In addition, fish and *Xenopus* are much less related to humans than for example rodents or other mammals, and thus will provide a more stringent screen for evolutionary conservation.

Acknowledgements

We thank Jack Leunissen (Center for Molecular and Biomolecular Informatics, Nijmegen, The Netherlands) for useful suggestions and K.J. Patel (Laboratory of Molecular Biology, Cambridge, UK) for critically reading the manuscript.

References

1 Auerbach AD, Buchwald M, Joenje H: Fanconi Anemia; in Scriver CR, Beadet AL, Valle D, Sly WS, Childs B, Kinzler KW, Vogelstein B (eds) (2001). The Metabolic and Molecular Basis of Inherited Disease, ed 8. New York, McGraw-Hill, pp 753–768.

2 Joenje H, Patel KJ (2001). The emerging genetic and molecular basis of Fanconi anaemia. *Nat Rev Genet* **2:** 446–457.

3 Sasaki MS, Tonomura A (1973). A high susceptibility of Fanconi's anemia to chromosome breakage by DNA cross-linking agents. *Cancer Res* **33:** 1829–1836.

4 Auerbach AD, Wolman SR (1976). Susceptibility of Fanconi's anaemia fibroblasts to chromosome damage by carcinogens. *Nature* **261:** 494–496.

5 Strathdee CA, Duncan AMV, Buchwald M (1992). Evidence for at least four Fanconi anaemia genes including FACC on chromosome 9. *Nat Genet* **1:** 196–198.

6 Joenje H, Oostra AB, Wijker M, di Summa FM, van Berkel CG, Rooimans MA, Ebell, W, van Weel M, Pronk JC, Buchwald M, Arwert F (1997). Evidence for at least eight Fanconi anemia genes. *Am J Hum Genet* **61:** 940–944.

7 Joenje H, Levitus M, Waisfisz Q, D'Andrea A, Garcia-Higuera I, Pearson T, van Berkel CG, Rooimans MA, Morgan N, Mathew CG, Arwert F (2000). Complementation analysis in Fanconi anemia: Assignment of the reference FA-H patient to group A. *Am J Hum Genet* **67:** 759–762.

8 Strathdee CA, Gavish H, Shannon WR, Buchwald M (1992). Cloning of cDNAs for Fanconi's anemia by functional complementation. *Nature* **356:** 763–767.

9 de Winter JP, Waisfisz Q, Rooimans MA, van Berkel CG, Bosnoyan-Collins L, Alon N, Carreau M, Bender O, Demuth I, Schindler D, Pronk JC, Arwert F, Hoehn H, Digweed M, Buchwald M, Joenje H (1998). The Fanconi anaemia group G gene is identical with XRCC9. *Nat Genet* **20:** 281–283.

10 de Winter JP, Leveille F, van Berkel CG, Rooimans MA, van der Weel L, Steltenpool J, Demuth I, Morgan NV, Alon N, Bosnoyan-Collins L, Lightfoot J, Leegwater PA, Waisfisz Q, Komatsu K, Arwert F, Pronk JC, Mathew CG, Digweed M, Buchwald M, Joenje H (2000). Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am J Hum Genet* **67:** 1306–1308.

11 de Winter JP, Rooimans MA, van der Weel L, van Berkel CG, Alon N, Bosnoyan-Collins L, de Groot J, Zhi Y, Waisfisz Q, Pronk JC, Arwert F, Mathew CG, Scheper RJ, Hoatlin ME, Buchwald M, Joenje H (2000). The Fanconi anemia gene FANCF encodes a novel protein with homology to ROM. *Nat Genet* **24:** 15–16.

12 Lo Ten Foe JR, Rooimans MA, Bosnoyan-Collins L, Alon N, Wijker M, Parker L, Lightfoot J, Carreau M, Callen DF, Savoia A, Cheng NC, van Berkel CG, Strunk MH, Gille JJ, Pals G, Kruyt FA, Pronk JC, Arwert F, Buchwald M, Joenje H (1996).

Expression cloning of a cDNA for the major Fanconi anaemia gene, *FAA*. *Nat Genet* **14:** 320–323.

13 The Fanconi Anaemia/Breast Cancer Consortium (1996). Positional cloning of the Fanconi anaemia group A gene. *Nat Genet* **14:** 324–328.

14 Timmers C, Taniguchi T, Hejna J, Reifsteck C, Lucas L, Bruun D, Thayer M, Cox B, Olson S, D'Andrea AD, Moses R, Grompe M (2001). Positional cloning of a novel Fanconi anemia gene, *FANCD2*. *Mol Cell* **7:** 241–248.

15 Kupfer GM, Naf D, Suliman A, Pulsipher M, D'Andrea AD (1997). The Fanconi anaemia proteins, FAA and FAC, interact to form a nuclear complex. *Nat Genet* **17:** 487–490.

16 Garcia-Higuera I, D'Andrea AD (1999). Regulated binding of the Fanconi anemia proteins FANCA and FANCC. *Blood* **93:** 1430–1432.

17 Waisfisz Q, de Winter JP, Kruyt FA, de Groot J, van der Weel L, Dijkmans LM, Zhi Y, Arwert F, Scheper RJ, Youssoufian H, Hoatlin ME, Joenje H (1999). A physical complex of the Fanconi anemia proteins FANCG/XRCC9 and FANCA. *Proc Natl Acad Sci USA* **96:** 10320–10325.

18 de Winter JP, van der Weel L, de Groot J, Stone S, Waisfisz Q, Arwert F, Scheper, RJ, Kruyt FA, Hoatlin ME, Joenje H (2000). The Fanconi anemia protein FANCF forms a nuclear complex with FANCA, FANCC and FANCG. *Hum Mol Genet* **9:** 2665–2674.

19 Yamashita T, Kupfer GM, Naf D, Suliman A, Joenje H, Asano S, D'Andrea AD (1998). The Fanconi anemia pathway requires FAA phosphorylation and FAA/FAC nuclear accumulation. *Proc Natl Acad Aci USA* **95:** 13085–13090.

20 Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J, Grompe M, D'Andrea AD (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7:** 249–262.

21 Li WH, Graur D (1991). Fundamentals of Molecular Evolution. Sunderland, Sinauer.

22 Lamb JR, Tugendreich S, Hieter P (1995). Tetratrico peptide repeat interactions: To TPR or not to TPR? *Trends Biochem Sci* **20:** 257–259.

23 Das AK, Cohen PW, Barford D (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: Implications for TPR-mediated protein-protein interactions. *EMBO J* **17:** 1192–1199.

24 Altschulz SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.

25 Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL (2000). The Pfam protein families database. *Nucleic Acids Res* **28:** 263–266.

26 van de Vrugt HJ, Cheng NC, de Vries Y, Rooimans MA, Groot J, Scheper RJ, Zhi Y, Hoatlin ME, Joenje H, Arwert F (2000). Cloning and characterization of murine Fanconi anemia group A gene: Fanca protein is expressed in lymphoid tissues, testis, and ovary. *Mamm Genome* **11:** 326–331.

27 Wong JC, Alon N, Norga K, Kruyt FA, Youssoufian H, Buchwald M (2000). Cloning and analysis of the mouse Fanconi anemia group A cDNA and an overlapping penta zinc finger cDNA. *Genomics* **67:** 273–283.

28 Wevrick R, Clarke CA, Buchwald M (1993). Cloning and analysis of the murine Fanconi anemia group C cDNA. *Hum Mol Genet* **2:** 655–662.

29 Nakanishi K, Moran A, Hays T, Kuang Y, Fox E, Garneau D, de Oca RM, Grompe M, D'Andrea AD (2001). Functional analysis of patient-derived mutations in the Fanconi anemia gene, FANCG/XRCC9. *Exp Hematol* **29:** 842–849

CHAPTER

3

Multiple TPR motifs characterize the Fanconi anemia FANCG protein

The genome protection pathway that is defective in patients with Fanconi anemia (FA) is controlled by at least eight genes, including *BRCA2*. A key step in the pathway involves the monoubiquitylation of FANCD2, which critically depends on a multi-subunit nuclear 'core complex' of at least six FANC proteins (FANCA, -C, -E, -F, -G, and -L). Except for FANCL, which has WD40 repeats and a RING finger domain, no significant domain structure has so far been recognized in any of the core complex proteins. By using a homology search strategy comparing the human FANCG protein sequence with its ortholog sequences in *Oryzias latipes* (Japanese rice fish) and *Danio rerio* (zebrafish) we identified at least seven tetratricopeptide repeat motifs (TPRs) covering a major part of this protein. TPRs are degenerate 34-amino acid repeat motifs which function as scaffolds mediating protein–protein interactions, often found in multiprotein complexes. In four out of five TPR motifs tested (TPR1, -2, -5, and -6), targeted missense mutagenesis disrupting the motifs at the critical position 8 of each TPR caused complete or partial loss of FANCG function. Loss of function was evident from failure of the mutant proteins to complement the cellular FA phenotype in FA-G lymphoblasts, which was correlated with loss of binding to FANCA. Although the TPR4 mutant fully complemented the cells, it showed a reduced interaction with FANCA, suggesting that this TPR may also be of functional importance. The recognition of FANCG as a typical TPR protein predicts this protein to play a key role in the assembly and/or stabilization of the nuclear FA protein core complex.

Eric Blom, Henri J. van de Vrugt, Yne de Vries,
Johan P. de Winter, Fré Arwert & Hans Joenje

DNA Repair 2004; 3: 77–84

Introduction

Fanconi anemia (FA) is a chromosomal instability syndrome characterized by developmental abnormalities, progressive pancytopenia and cancer-proneness [1]. The cellular phenotype is characterized by occurrence of spontaneous chromosomal aberrations and a hypersensitivity to agents that generate DNA crosslinks, such as mitomycin C (MMC) and diepoxybutane. Cell fusion studies revealed that FA is genetically heterogeneous with at least eleven complementation groups (A, B, C, D1, D2, E, F, G, I, J, and L) ([2–5]; Levitus et al., submitted for publication). At present, six FA genes have been identified by complementation cloning or by positional cloning techniques: *FANCA*, *-C*, *-D2*, *-E*, *-F*, and *-G* [2, 6–11], while a candidate-gene approach has led to identification of the breast cancer susceptibility gene *BRCA2* as the gene defective in FA-D1 patients [12]. The most recent FA gene discovery (*FANCL*) followed identification of a novel 43 kD WD40 repeat/RING finger protein that was associated with the nuclear FA protein core complex [5].

The nuclear core complex consists of at least six FA gene products (FANCA, -C, -E, -F, -G, and -L) [5, 13–17] and is required for the activation of FANCD2 by monoubiquitylation at Lys561 [18]. This modification is a key step in the pathway allowing further downstream reactions to take place, such as the association of FANCD2 with BRCA1 in nuclear repair foci. Both FANCD2 as well as FANCL are conserved in *Drosophila melanogaster*, but none of the other FANC proteins seem to exist in this organism [19, 20]. This suggests that at least a partial FA pathway is likely to be present in *Drosophila* and that the multi-protein core complex, found in vertebrates only, may have arisen later during evolution.

Here, we present the domain structure of FANCG. Although the core complex proteins FANCA, -C, -E, -F, and -G have been known for several years, no functional clues have yet been derived from their amino acid sequences. These proteins are only present in vertebrates and show no significant homology to other proteins, making their domain structure enigmatic. Previously, a putative peroxidase domain has been claimed in FANCA, but several amino acid substitutions that critically alter the consensus sequence had no effect on activity [21]. In addition, in FANCF a domain was reported that showed homology to the prokaryotic RNA binding protein ROM [9], but this homology is not preserved in the *Xenopus laevis* ortholog [Léveillé et al., manuscript in preparation] indicating that this motif is of no functional significance either. In order to elucidate the domain structure of FANCG, we identified and sequenced the zebrafish (*Danio rerio*) and the Japanese rice fish (*Oryzias latipes*) FANCG orthologs and analyzed them using a bioinformatic strategy. A preliminary report outlining the details of this approach has been published

previously [20]. In this study, we show that FANCG is composed of at least seven tetratricopeptide repeats (TPRs), together covering a major part of the protein. The TPR repeat is a highly degenerate 34-amino acid motif, typically present in tandem arrays of 3–16 motifs per protein. Each TPR motif is composed of a pair of anti-parallel α-helices of equal length and multiple TPRs fold the protein into a right-handed superhelical structure, which forms a scaffold for protein–protein interactions [22, 23]. Disturbance of individual TPRs by mutagenesis revealed that at least four TPRs are critical for the function of FANCG as assessed by the functional complementation of the cellular phenotype in FA-G cells. In addition, these TPRs are required for interaction with FANCA, confirming that they act as protein–protein interaction motifs. In conclusion, the presence of multiple TPRs in FANCG suggests an essential role for FANCG in the assembly and/or stabilization of the FA protein core complex.

Materials and methods

Constructs – An EST clone containing full-length *Danio rerio* (zebrafish) *fancg* cDNA was obtained from RZPD (Deutsches Ressourcenzentrum für Genomforschung, Germany, clone IMAGp998A239112Q2) and sequenced (GenBank acc. no. AJ496390). An *Oryzias latipes* (Japanese rice fish) EST clone containing full-length *fancg* cDNA (clone name MF01SSA194A05) was a kind gift of Yumiko Saga (National Institute of Genetics, Japan) and was also sequenced (GenBank acc. no. AJ566763). *FANCG* mutant constructs containing HA-tagged human *FANCG* cDNA in the episomal mammalian expression vector pMEP4 (Invitrogen) were generated using standard PCR mutagenesis methods. All mutant PCR fragments were amplified using Platinum Pfx DNA proofreading polymerase (Invitrogen) and sequenced in full-length to confirm the presence of the intended mutation and the absence of other mutations.

Bioinformatic analysis – Danio rerio and *Oryzias latipes* fancg protein sequences were analyzed as previously described [20] and as outlined in the results section.

Cell culture and transfection – EUFA316 (FA-G) lymphoblastoid cells were grown in RPMI 1640 medium (Gibco-BRL) supplemented with 10% fetal bovine serum (Gibco-BRL) at 37°C under an atmosphere of 5% CO2 in air. To stably transfect cells, 4×10^6 exponentially growing cells were electroporated (BTX Electro Square Porator ECM830) with 10 μg plasmid DNA (QIAfilter Plasmid Midi kit, Qiagen) and the transfected cells were then selected for 3 weeks in medium supplemented with 100

μg/ml Hygromycin B (Roche Molecular Biochemicals). Cells carrying the pMEP4 vector were grown in the presence of 40 μM Zn^{2+} to induce transcription from the metallothionein promoter. This concentration of Zn^{2+} had no effect on cell growth.

MMC-induced growth inhibition test – To test functional complementation of the TPR-mutated *FANCG* constructs, stably transfected EUFA316 cells were exposed to 0, 1, 2, 3, 6, 10, 30, 60, and 100 nM mitomycin C (Kyowa Hakko Kyogo Co., Ltd., Japan) in parallel cultures for several days (approximately three population doublings), after which the percentage growth was determined using a Coulter counter, as described previously [24].

Coimmunoprecipitation and immunoblotting – Total cell extracts were prepared by lysis of 5×10^6 cells in 0.5 ml lysis buffer (50 mM Tris–HCl pH7.4, 150 mM NaCl, 1% Nonidet P40, 500 μg/ml Pefabloc, 1 μg/ml apoprotinin, 1 μg/ml pepstatin, 1 μg/ml leupeptin) for 10 min, on ice. Lysates were sheared through a 19-gauge needle, then clarified by centrifugation (14,000 rpm, 4°C) in a microcentrifuge, after which 2 μl of α-HA antibody (12CA5 mouse monoclonal, Boehringer Mannheim) was added and incubated at 4°C, overnight. Antibody-bound proteins were collected with 40 μl protein A agarose beads (Invitrogen) for 30 min at 4°C and washed three times in lysis buffer, after which samples were analyzed by immunoblotting.

For immunoblotting, 20 μl of 2× sample buffer (30 mM Tris–HCl pH 6.8, 0.8% sodium dodecyl sulfate (SDS), 5% 2-mercaptoethanol, 4% glycerol, 0.4% bromphenol blue) was added and samples were boiled for 5 min. After centrifugation, the supernatant was subjected to standard 8% polyacrylamide SDS gel electrophoresis after which proteins were transferred to an Immobilon-P transfer membrane (Millipore) in transfer buffer (25 mM Tris–HCl, 200 mM glycine, 15% methanol) at 300 mA, at 4 °C for 2 h. The membrane was blocked for 30 min in 5% non-fat dry milk in TBST (10 mM Tris–HCl, 150 mM NaCl, 0.1% Tween-20), washed in TBST, and incubated with primary antibody (α-HA 12CA5 or α-FANCA Rb89) in 2% non-fat dry milk/TBST, overnight at 4°C. After washing three times in TBST, the membrane was incubated in secondary antibody (goat anti-mouse-HRP or goat anti-rabbit, DAKO, Denmark) in 2% non-fat dry milk/TBST for 1 h, at room temperature. The membrane was then washed five times in TBST and developed by using the ECL Western blotting analysis system (Amersham Pharmacia) according to manufacturer's instructions.

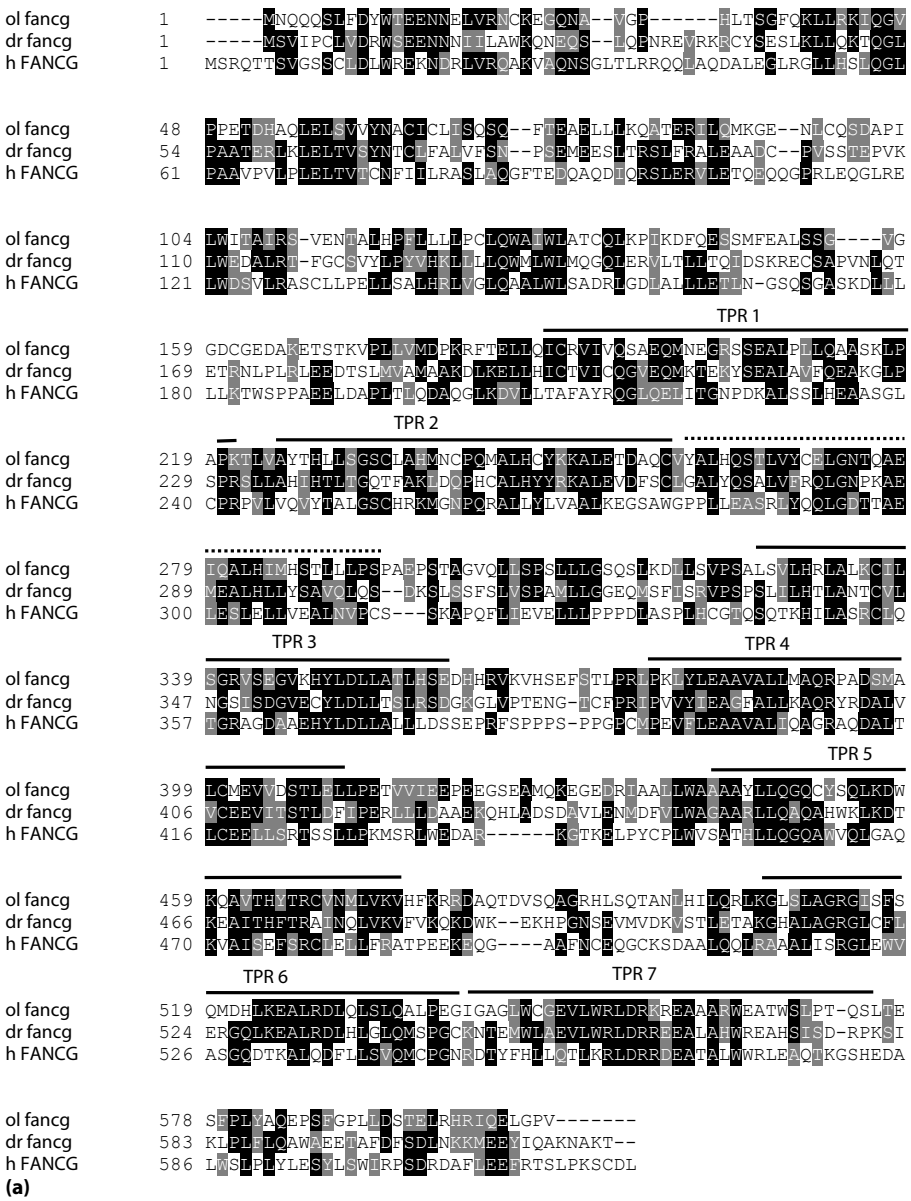


Fig 1. Comparative sequence analysis of *Oryzias latipes* (Japanese rice fish) and *Danio rerio* (zebrafish) orthologs with human FANCG. (a) ClustalW alignment of *Oryzias latipes*, *Danio rerio* and human FANCG. Identical residues are indicated in black and similar residues in gray (BoxShade program). The regions containing the TPR motifs are indicated on top of the sequence. A putative additional TPR, with a low homology to the TPR consensus sequence, is indicated with a dotted line. (b) Comparison of the 7 identified TPRs of FANCG.

Fig 1. *(Continued).*

Results

FANCG is a member of the family of TPR-containing proteins – We used a computational approach in our search for possible clues about the domain structure of FANCG. To ensure maximum sensitivity of the analysis, we compared the human protein with that of the zebrafish and with that of the Japanese rice fish, which are both distant orthologs of the human FANCG sequence (Fig. 1a). Orthologs are genes found in different species that have originated from a common ancestor gene at an earlier evolutionary time. They help to identify conserved residues or regions which may be functionally important. Human and zebrafish show 26% identity and 41% similarity, and human and Japanese rice fish show 24% identity and 39% similarity. In addition, the two fish sequences share 34% identity and 52% similarity. Using the two fish orthologs and the PSI-BLAST algorithm [25], distant but statistically significant homologies were found between FANCG and several TPR-containing proteins, such as the O-GlcNAc transferase p110 subunit. To confirm the possible presence of TPRs in FANCG, further analysis was then performed with the Pfam [26] and REP [27] programs, as a result of which a total of seven TPR motifs were identified (Fig. 1a). Moreover, region 281–314 (between TPR2 and TPR3) is likely to contain an additional TPR, but showed less homology to the TPR consensus sequence. Using a profile search strategy, based on the TPRs from FANCG (Fig. 1b), the presence of these motifs in other TPR proteins (such as, cell division control protein cdc27, O-GlcNAc transferase p110 subunit, translocase of outer membrane TOM70, TPR repeat protein 2, and stress-induced-phosphoprotein 1 STI1) could correctly be predicted (HMMER and MAST algorithms). This further confirms the identified motifs as TPRs.

Within the seven TPRs, only a few amino acids were found to be conserved in all (Fig. 1b). Position 8 contains in most cases the small and hydrophobic residues Gly or Ala, whereas positions 20–21 usually contain Ala-Leu. The TPR motif folds into two anti-parallel α-helices and these consensus residues are all located at the interface between the two helices, forming a part of the hydrophobic core of the motif [23]. Like in other TPR proteins, positions that are outside the consensus residues are far less conserved.

TPR1, -2, -5 and -6 are critical to the function of FANCG – To investigate the functional significance of the TPRs, we mutated the conserved 8th position (Gly or Ala) into the more bulky Gln to disrupt the individual TPRs. Functional activity was then determined by the ability of the constructs to correct the mitomycin C (MMC) sensitivity of the FA-G lymphoblastoid cell line EUFA316. All mutants were cloned in the pMEP4 vector and cells were therefore grown in the presence of 40 μMZn^{2+} to induce transcription from its metallothionein promoter. TPR1 (G216Q) showed an intermediate complementation, but TPR5 (G460Q) and TPR6 (G521Q) showed no complementation (Fig. 2a), indicating that these three TPRs are required for the FA pathway to function. In contrast, TPR4 (A401Q) showed a wild type-like phenotype, suggesting that this motif may not be functionally important. Since the human FANCG sequence contains no Gly or Ala at position 8 in TPR7 (Fig. 1b), another missense mutation was chosen that was not at a consensus position (R563E), but this mutant behaved as wild type making the results for this mutant inconclusive. In addition, TPR3 was identified only recently after the complementation studies had been completed and therefore this mutant has not been tested to date. TPR3 was found using a profile search based on the six previously identified TPRs from all three species (MAST algorithm).

Of the mutants studied, only TPR2 (G253Q) showed expression level-dependent complementation (Fig. 2b). When cells were not grown in the presence of Zn^{2+}, the promoter activity of the metallothionein promoter is known to be 'leaky' resulting in a much lower, but sufficient expression for full complementation by the wild type, TPR4 mutant, and TPR7 mutant constructs (Fig. 2b, data not shown for TPR4 and TPR7). However, this was not the case for the TPR2 mutant. This mutant showed a clear sensitivity to MMC when the promoter was not induced. Since the uninduced condition may resemble the normal endogenous expression more closely, we speculate that this mutation might be pathogenic in vivo. These results show that at least four out of seven TPRs in FANCG are important for a functional FA pathway.

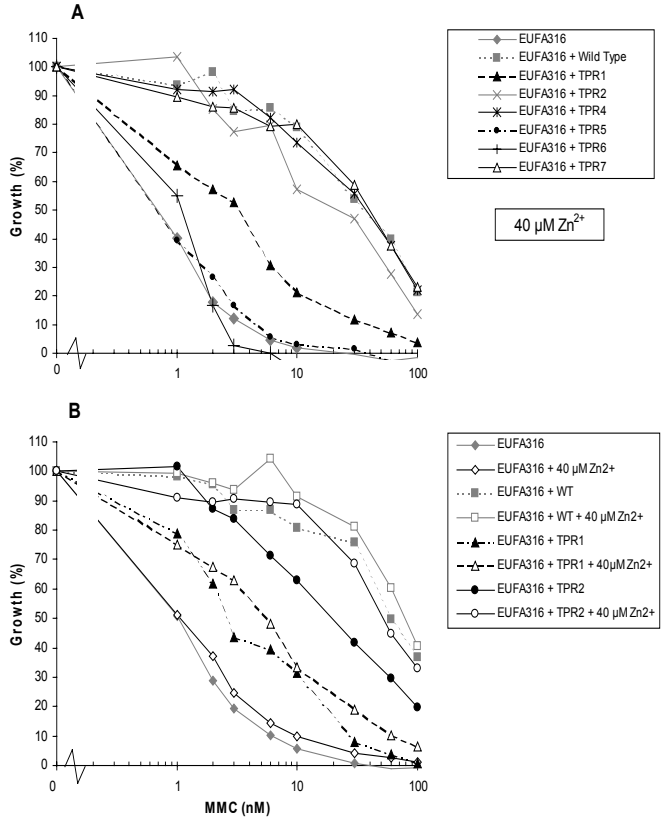


Fig 2. Correction of the cellular FA phenotype by TPR-mutated constructs. (a) All constructs were in pMEP4 vector and cells were, therefore, cultured in presence of 40 µM Zn^{2+} to induce transcription from the metallothionein promoter. (b) Expression level-dependent complementation of FA-G cells by the TPR2 mutant, but not by other TPR-mutated constructs (data not shown for TPR4-7). Expression of constructs was verified by Western blotting (data not shown).

TPR1, -2, -4, -5, and -6 are important for interaction with FANCA – Since the TPR motif is known to be a protein–protein interaction motif, typically present in large protein complexes, we tested the ability of the TPR-mutants to interact with FANCA, which is a known direct interaction [28, 29]. TPR1, -2, -5, and -6 mutants showed no detectable precipitation of FANCA with FANCG (Fig. 3), indicating that a large region in FANCG is required for this interaction. Although TPR4 mutant still complemented FA-G cells, the affinity for FANCA appeared to be reduced, suggesting that this TPR may also be involved in binding to FANCA. The loss of interaction with FANCA confirms that the TPR motifs in FANCG, like in other TPR-containing proteins, function as protein–protein interaction motifs.

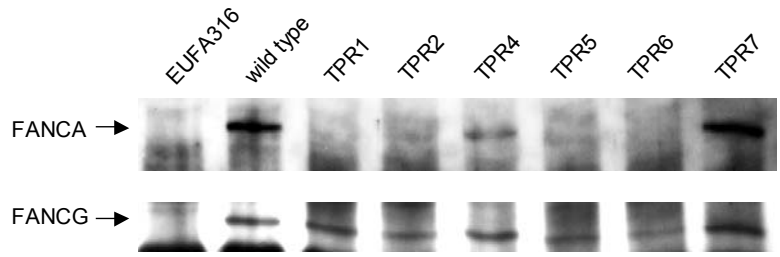


Fig 3. Co-immunoprecipitation of FANCG and FANCA. EUFA316 cells were stably transfected with the indicated HA-tagged pMEP4 expression constructs and cells were treated with 40 μM Zn^{2+} to induce transcription from the metallothionein promoter of the vector. Cell extracts were analyzed by immunoprecipitation with α-HA antibody followed by Western blotting with α-HA and α-FANCA antibodies.

Discussion

This study reveals that FANCG is a member of the family of TPR-containing proteins. We identified seven TPR motifs covering the majority of the protein. The tetratricopeptide repeat is a highly degenerate 34-amino acid motif typically present in tandem arrays of 3–16 motifs per protein. Each TPR is composed of a pair of anti-parallel α-helices of equal length and the multiple TPRs fold the protein into a right-handed superhelical structure [30]. This structure forms a scaffold for protein–protein interaction that is thought to be essential for the assembly of multiprotein complexes. TPRs have been described in a wide variety of proteins including the anaphase-promoting complex subunits cdc16, cdc23, and cdc27; hsp-90-binding immunophilins, peroxisomal and mitochondrial import proteins; the PKR kinase inhibitor; Strap, a cofactor in the p300 coactivator complex, and XAB2, a protein involved in transcription-coupled DNA repair and transcription [23, 30–32]. Since TPR proteins are typically present in large protein complexes, the presence of TPRs in FANCG strongly suggests that FANCG has a role in the assembly and/or stabilization of the nuclear FA protein core complex. This is in agreement with a recent three-hybrid study which suggested that FANCG mediates an interaction between FANCA and FANCF by forming a physical bridge between these proteins [33].

We used a mutagenesis approach to assess the functional importance of the TPRs by creating mutations at their positions 8. Mutations at this consensus position have been reported to result in a disruption of their function in several other TPR-containing proteins [34–36]. All mutants that were mutated at their 8th position, except the TPR4 mutant, were unable to correct the cellular defect in FA-G cells. This suggests that this consensus position may also be a reporter of function in the FANCG protein. The slightly reduced interaction of the TPR4 mutant with FANCA

and the high conservation of the TPR4 region between the two fishes and the human sequence both indicate that TPR4 may be functionally relevant as well. Although the human TPR7 sequence contains Leu at its 8th residue and the R563E mutant complemented FA-G cells as wild type, we speculate that this TPR may still be a 'genuine' TPR. The zebrafish and the Japanese rice fish sequences show strong homology to the TPR consensus, while between the two fishes and the human sequence many residues are conserved. Further studies with additional mutants for TPR4 and TPR7 may help to evaluate the functional relevance for these two TPRs.

Our results show that most TPRs in FANCG are required for interaction with FANCA, verifying these TPRs as interaction domains and indicating that a region spanning at least the residues 209–581 is required for interaction with FANCA. The interaction between FANCG and FANCA has previously been described as a direct interaction involving amino acids 18-29 of FANCA [28, 29]. Moreover, several FANCG deletion studies, all with different results, have been reported. Gordon and Buchwald recently showed by yeast two-hybrid technique that only a fragment with a C-terminal deletion of 142 residues is able to interact with FANCA and that longer C-terminal deletions are not able to interact [33]. In another yeast two-hybrid study [28], fragments 1-367, 305-622, and 511-622 are all reported to interact with FANCA, and this is in apparent contradiction with the study of Gordon and Buchwald. In addition, Kruyt et al. [29] reported that in vitro translated 1-585, 1-550, and 1-475 fragments are all able to interact weakly by coimmunoprecipitation but not a fragment spanning 1-400. In tertiary structure, TPRs are known to form a superhelical structure which forms a groove that mediates interaction with other proteins. We speculate that fragments of FANCG, as used in these deletion studies, may only consist of a partial superhelix and may therefore have a reduced affinity for their substrate. Weak interactions may then only be detected by sensitive techniques such as the yeast two-hybrid system.

In conclusion, the FANCG protein appears to lose its previous qualification as an 'orphan'. At present, FANCG and the very recently identified FANCL protein are the only 'upstream' FANC proteins that have been characterized in terms of their domain structure. Identification and sequencing of the two fish orthologs have greatly facilitated the recognition of all TPRs in FANCG. With merely the human sequence available only three TPRs would have been recognized. In further studies, the cloning and sequencing of vertebrate orthologs of the other FANC proteins may, in a similar way, help to elucidate their molecular function.

Acknowledgements

We would like to thank Ina Rolfs (RZPD Deutsches Ressourcenzentrum für Genomforschung, Germany) for help with the zebrafish EST clone, Yumiko Saga (National Institute of Genetics, Japan) for the kind gift of the Japanese rice fish EST clone, and Annette Medhurst (VU University Medical Center, Amsterdam) for critically reading the manuscript.

References

1 D'Andrea AD, Grompe M (2003). The Fanconi anaemia/BRCA pathway. *Nat Rev Cancer* **3:** 23–34.

2 Timmers C, Taniguchi T, Hejna J, Reifsteck C, Lucas L, Bruun D, Thayer M, Cox B, Olson S, D'Andrea AD, Moses R, Grompe M (2001). Positional cloning of a novel Fanconi anemia gene, *FANCD2*. *Mol Cell* **7:** 241–248.

3 Joenje H, Levitus M, Waisfisz Q, D'Andrea AD, Garcia-Higuera I, Pearson T, van Berkel CG, Rooimans MA, Morgan N, Mathew CG, Arwert F (2000). Complementation analysis in Fanconi anemia: assignment of the reference FA-H patient to group A. *Am J Hum Genet* **67:** 759–762.

4 Joenje H, Oostra AB, Wijker M, di Summa FM, van Berkel CG, Rooimans MA, Ebell W, van Weel M, Pronk JC, Buchwald M, Arwert F (1997). Evidence for at least eight Fanconi anemia genes. *Am J Hum Genet* **61:** 940–944.

5 Meetei AR, de Winter JP, Medhurst AL, Wallisch M, Waisfisz Q, van de Vrugt HJ, Oostra AB, Yan Z, Ling C, Bishop CE, Hoatlin ME, Joenje H, Wang W (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35:** 165–170.

6 The Fanconi anaemia/breast cancer consortium (1996). Positional cloning of the Fanconi anaemia group A gene. *Nat Genet* **14:** 324–328.

7 Lo Ten Foe JR, Rooimans MA, Bosnoyan-Collins L, Alon N, Wijker M, Parker L, Lightfoot J, Carreau M, Callen DF, Savoia A, Cheng NC, van Berkel CG, Strunk MH, Gille JJ, Pals G, Kruyt FA, Pronk JC, Arwert F, Buchwald M, Joenje H (1996). Expression cloning of a cDNA for the major Fanconi anaemia gene, *FAA*. *Nat Genet* **14:** 320–323.

8 Strathdee CA, Gavish H, Shannon WR, Buchwald M (1992). Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **356:** 763–767.

9 de Winter JP, Rooimans MA, van der Weel L, van Berkel CG, Alon N, Bosnoyan-Collins L, de Groot J, Zhi Y, Waisfisz Q, Pronk JC, Arwert F, Mathew CG, Scheper RJ, Hoatlin ME, Buchwald M, Joenje H (2000). The Fanconi anaemia gene *FANCF* encodes a novel protein with homology to ROM. *Nat Genet* **24:** 15–16.

10 de Winter JP, Léveillé F, van Berkel CG, Rooimans MA, van der Weel L, Steltenpool J, Demuth I, Morgan NV, Alon N, Bosnoyan-Collins L, Lightfoot J, Leegwater PA, Waisfisz Q, Komatsu K, Arwert F, Pronk JC, Mathew CG, Digweed M, Buchwald M, Joenje H (2000). Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am J Hum Genet* **67:** 1306–1308.

11 de Winter JP, Waisfisz Q, Rooimans MA, van Berkel CG, Bosnoyan-Collins L, Alon N, Carreau M, Bender O, Demuth I, Schindler D, Pronk JC, Arwert F, Hoehn H, Digweed M, Buckwald M, Joenje H (1998). The Fanconi anaemia group G gene *FANCG* is identical with *XRCC9*. *Nat Genet* **20:** 281–283.

12 Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C, Persky N, Grompe M, Joenje H, Pals G, Ikeda H, Fox EA, D'Andrea AD (2002). Biallelic inactivation of *BRCA2* in Fanconi anemia. *Science* **297:** 606–609.
13 Kupfer GM, Naf N, Suliman A, Pulsipher M, D'Andrea AD (1997). The Fanconi anaemia proteins, FAA and FAC, interact to form a nuclear complex. *Nat Genet* **17:** 487–490.
14 Garcia-Higuera I, D'Andrea AD (1999). Regulated binding of the Fanconi anemia proteins, FANCA and FANCC. *Blood* **93:** 1430–1432.
15 Waisfisz Q, de Winter JP, Kruyt FA, de Groot J, van der Weel L, Dijkmans LM, Zhi Y, Arwert F, Scheper RJ, Youssoufian H, Hoatlin ME, Joenje H (1999). A physical complex of the Fanconi anemia proteins FANCG/XRCC9 and FANCA. *Proc Natl Acad Sci USA* **96:** 10320–10325.
16 de Winter JP, van der Weel L, de Groot J, Stone S, Waisfisz Q, Arwert F, Scheper RJ, Kruyt FA, Hoatlin ME, Joenje H (2000). The Fanconi anemia protein FANCF forms a nuclear complex with FANCA, FANCC and FANCG. *Hum Mol Genet* **9:** 2665–2674.
17 Pace P, Johnson M, Tan WM, Mosedale G, Sng C, Hoatlin M, de Winter J, Joenje H, Gergely F, Patel KJ (2002). FANCE: the link between Fanconi anaemia complex assembly and activity. *EMBO J* **21:** 3414–3423.
18 Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J, Grompe M, D'Andrea AD (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7:** 249–262.
19 Castillo V, Cabré O, Marcos R, Surrallés J (2003). Molecular cloning of the *Drosophila* Fanconi anaemia gene *FANCD2* cDNA. *DNA Rep. (Amst.)* **2:** 751–758.
20 Blom E, van de Vrugt HJ, de Winter JP, Arwert F, Joenje H (2002). Evolutionary clues to the molecular function of Fanconi anemia genes. *Acta Haematol* **108:** 231–236.
21 Ren J, Youssoufian H (2001). Functional analysis of the putative peroxidase domain of FANCA the Fanconi anemia complementation group A protein. *Mol Genet Metab* **72:** 54–60.
22 Groves MR, Barford D (1999). Topological characteristics of helical repeat proteins. *Curr Opin Struct Biol* **9:** 383–389.
23 Lamb JR, Tugendreich S, Hieter P (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* **20:** 257–259.
24 Joenje H, Lo ten Foe JR, Oostra AB, van Berkel CG, Rooimans MA, Schroeder-Kurth T, Wegner RD, Gille JJ, Buchwald M, Arwert F (1995). Classification of Fanconi anemia patients by complementation analysis: evidence for a fifth genetic subtype. *Blood* **86:** 2156–2160.

25 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25:** 3389–3402.
26 Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002). The Pfam protein families database. *Nucl Acids Res* **30:** 276–280.
27 Andrade MA, Ponting CP, Gibson TJ, Bork P (2000). Homology-based method for identification of protein repeats using statistical significance estimates. *J Mol Biol* **298:** 521–537.
28 Huber PA, Medhurst AL, Youssoufian H, Mathew CG (2000). Investigation of Fanconi anemia protein interactions by yeast two-hybrid analysis. *Biochem Biophys Res Commun* **268:** 73–77.
29 Kruyt FA, Abou-Zahr F, Mok H, Youssoufian H (1999). Resistance to mitomycin C requires direct interaction between the Fanconi anemia proteins FANCA and FANCG in the nucleus through an arginine-rich domain. *J Biol Chem* **274:** 34212–34218.
30 Das AK, Cohen PW, Barford D (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein–protein interactions. *EMBO J* **17:** 1192–1199.
31 Demonacos C, Krstic-Demonacos M, La Thangue NB (2001). A TPR motif cofactor contributes to p300 activity in the p53 response. *Mol Cell* **8:** 71–84.
32 Nakatsu Y, Asahina H, Citterio E, Rademakers S, Vermeulen W, Kamiuchi S, Yeo JP, Khaw MC, Saijo M, Kodo N, Matsuda T, Hoeijmakers JH, Tanaka K (2000). XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *J Biol Chem* **275:** 34931–34937.
33 Gordon SM, Buchwald M (2003). The Fanconi anemia protein complex: mapping protein interactions in the yeast two- and three-hybrid systems. *Blood* **102:** 136–141.
34 Gounalaki N, Tzamarias D, Vlassi M (2000). Identification of residues in the TPR domain of Ssn6 responsible for interaction with the Tup1 protein. *FEBS Lett* **473:** 37–41.
35 Sikorski RS, Michaud WA, Hieter P (1993). p62cdc23 of *Saccharomyces cerevisiae*: a nuclear tetratricopeptide repeat protein with two mutable domains, *Mol Cell Biol* **13:** 1212–1221.
36 Koga H, Terasawa H, Nunoi H, Takeshige K, Inagaki F, Sumimoto H (1999). Tetratricopeptide repeat (TPR) motifs of p67(phox) participate in interaction with the small GTPase Rac and activation of the phagocyte NADPH oxidase. *J Biol Chem* **274:** 25051–25060.

CHAPTER

4

The Fanconi anemia gene product FANCF is a flexible adaptor protein

The Fanconi anemia (FA) protein FANCF is an essential component of a nuclear core complex that protects the genome against chromosomal instability, but the specific function of FANCF is still poorly understood. Based upon the homology between human and *Xenopus laevis* FANCF, we carried out an extensive mutagenesis study to examine which domains are functionally important and to gain more insight into the function of FANCF. In contrast to previous suggestions, we show that FANCF does not have a ROM-like function. We found that the C-terminus of FANCF interacts directly with FANCG and allows the assembly of other FA proteins into a stable complex. The N-terminus appears to stabilize the interaction with FANCA and FANCG, and is essential for the binding of the FANCC/FANCE subcomplex. We identified several important amino acids in this N-terminal region but, surprisingly, many amino acid changes failed to affect the function of the FANCF protein. Our data demonstrate that FANCF acts as a flexible adaptor protein that plays a key role in the proper assembly of the FA core complex.

France Léveillé, Eric Blom, Annette L. Medhurst, Patrick Bier, El Houari Laghmani, Mark Johnson, Martin A. Rooimans, Alexandra Sobeck, Quinten Waisfisz, Fré Arwert, K. J. Patel, Maureen E. Hoatlin, Hans Joenje & Johan P. de Winter

J Biol Chem 2004; 279: 39421–39430

Introduction

Fanconi anemia (FA) is an autosomal recessive chromosomal instability syndrome. The clinical phenotype of FA patients is characterized by congenital abnormalities, progressive bone marrow failure, and a predisposition to cancer; particularly acute myeloid leukemia and squamous cell carcinoma [1, 2]. The spontaneous cytogenetic aberrations specific for FA cells are exacerbated upon treatment with DNA cross-linking agents, such as mitomycin C (MMC) and diepoxybutane, which suggests a DNA maintenance defect particularly in the handling of cross-link damage.

To date, somatic cell fusion studies have demonstrated 11 FA complementation groups (A–C, D1, D2, E–G, I, J, L) [3] and eight of the FA associated genes have been identified: *FANCA*, *FANCC*, *FANCD1/BRCA2, FANCD2*, *FANCE*, *FANCF*, *FANCG* and *FANCL* [4–12]. These FA genes encode 'orphan proteins' and no functional domains have been found in their primary amino acid sequences, except for a RING finger in FANCL [10] and multiple TPR motifs in FANCG [13]. In FANCF a region with homology to RNA-binding protein ROM has been suggested to provide a clue for its function [5].

FANCA, FANCC, FANCE, FANCF, FANCG and FANCL assemble in a nuclear core complex [10, 14–17], in which FANCG binds directly to FANCA and FANCF and FANCC binds directly to FANCE [15, 16]. The nuclear core complex is essential for the monoubiquitination of FANCD2 [18] and this modified form of FANCD2 colocalizes with BRCA1, RAD51 [19] and PCNA [20] in foci that also contain other DNA repair proteins. Nevertheless, it is still unknown how monoubiquitinated FANCD2 is involved in DNA repair and MMC resistance. Recently, biallelic mutations in *BRCA2* have been found in cell lines derived from FA-D1 patients, adding BRCA2 to the list of FA proteins [8]. The BRCA proteins are known to be involved in a multitude of biological functions including DNA repair, recombination, cell cycle control and transcription [21]. BRCA2 appears to be directly linked to the repair of double-stranded breaks (DSBs) by homologous recombination [22]. FANCD2 has also been identified as a target of the ataxia telangiectasia (AT) signaling pathway [23] and seems to be functionally connected to the RAD50/MRE11/NBS1 (RMN) protein complex [24, 25], which plays an important role in the repair of DSBs [26]. Taken together, these data suggest that the FA proteins serve to maintain genomic stability and integrity, in concert with other protein complexes.

FANCF plays an important role in the FA pathway and its functional disruption seems to be involved in specific types of cancer, as suggested from the hypermethylation of the *FANCF* promoter in a subset of ovarian, oral, lung and cervical cancers [27–29]. Nevertheless, the function of FANCF is still poorly

understood. Patient-derived mutant forms of FA proteins and structure/function analysis of FANCA, FANCC and FANCG have been informative in finding important functional residues [30–34], but for FANCF this information is lacking, since none of the FA-F patients have missense mutations that could provide insight into its function [5]. For this reason, we started a site-directed mutagenesis study to obtain more information about functional domains in the FANCF protein. Because such domains are expected to be relatively conserved during evolution, we searched for FANCF homologs in lower vertebrates that might highlight important residues in FANCF. We found a *Xenopus laevis* homolog of FANCF (xFANCF), which has a relatively low overall homology with human FANCF. However, two relatively conserved regions were located at the N- and C- terminus. Based upon this conservation we generated a large panel of FANCF mutants and identified several functionally important amino acids and domains in FANCF. This study reveals that FANCF is an adaptor protein that plays a key role in the proper assembly of the FA core complex. To be able to perform this function the N-terminus of FANCF interacts with the FANCC/FANCE subcomplex, while the C-terminus binds to the FANCA/FANCG subcomplex.

Material and methods

Cell culture and transfection – Epstein-Barr virus (EBV)-transformed lymphoblasts were cultured in RPMI 1640 media supplemented with 1 mM glutamine (Gibco, Invitrogen, Carlsbad, CA, USA) and 10% fetal calf serum (FCS; Gibco). Selection medium to obtain stable cell lines also contained hygromycin B (100 µg/ml; Roche, Basel, Switzerland). For stable expression lymphoblastoid cell lines were transfected by electroporation using an ECM830 electro square porator (BTX, San Diego, CA, USA). The MMC-induced growth inhibition assays were performed as previously described [35, 36].

Generation of FANCF mutant constructs – The FANCF mutant constructs were generated by polymerase chain reaction (PCR) with oligonucleotides encoding the amino acid substitutions or deleted regions of the FANCF sequence. We used as a template the FANCF cDNA clone 10 obtained by expression cloning [5]. The FANCF mutant R47A+F48A was made using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA) according to the manufacturer's instructions. All FANCF mutants have been analyzed by sequencing to confirm that the desired changes had occurred and to exclude the presence of other PCR derived alterations. The mutants were subcloned into the expression vector pCEP4 (Invitrogen) and transfected in FA-F lymphoblasts, EUFA698. FANCF mutants 1-15del and delC31,

and wild type FANCF-Flag were also subcloned into the expression vector pIRESneo (Clontech, Palo Alto, CA, USA) to generate stable cell lines. Expression of the mutant proteins in the stable cell lines was confirmed by immunoprecipitation and immunoblotting with FANCF specific antibodies.

Sequence analysis – FANCF mutants and the *Xenopus laevis* FANCF IMAGE clone 3200942 (obtained from HGMP Resource Centre, Cambridge, UK), were subcloned in pBluescript SK^- and were sequenced with CY5.5 labeled T7 and T3 primers using a Thermo Sequenase primer cycle sequencing kit (Amersham). Products were analyzed on a Visible Genetics automatic DNA sequencer (Visible Genetics Inc, Toronto, Ontario, Canada).

Yeast two-hybrid analysis – The MATCHMAKER Two-Hybrid System 3 (Clontech) was used according to the manufacturer's instructions as previously described [15]. In brief, bait and prey constructs were sequentially transformed into AH109 yeast cells and selected on -trp-leu-his-ade medium. The resulting colonies were tested for β-galactosidase expression with X-gal. To confirm interactions a yeast-mating assay was used where constructs were transformed separately into two different yeast strains and mating cultures plated onto selection medium as above. All constructs were tested for self-activation against a series of control plasmids and expression of the mutant proteins was verified by immunoblotting.

Mammalian two- and three-hybrid analysis – Human embryonic kidney cells 293, either untransfected or stably transfected with wild type or mutant (L554P) FANCC were plated onto six well plates. After 48h, the cells were transiently transfected with FANCE cDNA fused to the GAL-4 activation domain (pVP16; Clontech) and the indicated FANCF constructs fused in frame to the GAL4 DNA-binding domain (pM; Clontech) (1 μg of each), together with a GAL4 driven reporter plasmid (G5E1bLUC, 0.2 μg). The luciferase activity was monitored after 24 h using a Dual-Luciferase Reporter Assay System (Promega) and a single tube luminometer (DLReady, Berthold Detection Systems), according to the manufacturer's instructions. All GAL4 constructs were sequenced to confirm the correct reading frame and each experimental data set was performed in triplicate to overcome the variability inherent to transfections. Transient expression of the FANCF constructs and stable expression of FANCC were confirmed by immunoblotting with FANCF and FANCC specific antisera.

Subcellular fractionation, immunoprecipitation and immunoblotting – Nuclear and cytoplasmic fractions of the lymphoblastoid cell lines were obtained as described previously [17]. Cell extracts (~10^7 cells) were prepared in lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl and 1% Nonidet NP40 supplemented with protease inhibitors) and subjected to immunoprecipitation using the indicated antibodies, as previously described (14). Anti-Flag M2 affinity gel (Sigma-Aldrich, Saint-Louis, MO, USA) was used to immunoprecipitate Flag-tagged proteins. Immunoprecipitates or lysates were then separated on 8% SDS-polyacrylamide gels, transferred to PVDF membranes and specific proteins were detected by immunoblotting with the indicated antibodies.

Results

Homology between the human and Xenopus FANCF proteins – Since functionally important amino acids are well conserved between species, we searched several databases for FANCF homologs to obtain clues about functional domains. In the NCBI database an IMAGE clone (3200942) was found, which represented the full-length *Xenopus laevis* homolog of *FANCF* (*xFANCF*). The cDNA was sequenced and the predicted protein was aligned with the human FANCF sequence (Fig. 1). The overall homology between both sequences appeared to be limited (27% identity, 49% similarity) and dispersed over the whole molecule. However, relatively high sequence conservation was found in both the N- and C-terminal regions of the protein. To test if this limited homology was sufficient to function in a human background, Flag-tagged xFANCF was transfected into a cell line from an FA-F patient (EUFA698). Unlike the human protein, xFANCF was unable to complement the MMC hypersensitivity of the cells, despite of its proper expression (Fig. 2). The untagged xFANCF protein was also unable to complement the MMC defect in FA-F cells (data not shown), indicating that xFANCF is not able to restore the human FA pathway.

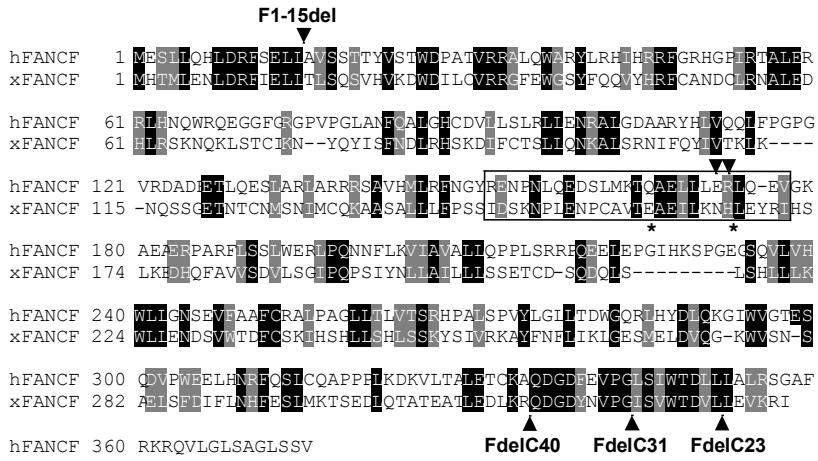


Fig. 1. Sequence alignment of human and *Xenopus laevis* FANCF proteins. Alignments were made with BoxShade. The black-shaded residues are identical and the gray-shaded residues are similar between human and *Xenopus* FANCF. The box shows the region homologous to the prokaryotic RNA-binding protein ROM. The asterisks point out the amino acids essential for RNA-binding activity in ROM and the triangles indicate the amino acids that were mutated in the ROM homologous region. The N- and C-terminal deletion mutants that were generated are also indicated. GenBank Accession Number for xFANCF: AY547288.

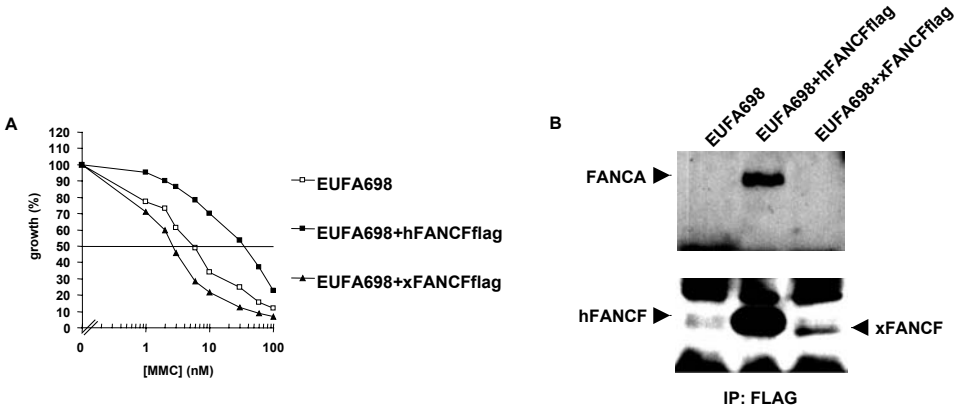


Fig. 2. *Xenopus* FANCF protein is not functional in human cells**.** (A) The MMC hypersensitive phenotype of the FA-F lymphoblastoid cell line EUFA698 is corrected by stable transfection with cDNA encoding human FANCF, but not with cDNA encoding *Xenopus* FANCF. (B) Expression of Flag-tagged human and *Xenopus* FANCF proteins in EUFA698 cells. Cell lysates of the indicated cell lines were immunoprecipitated with anti-flag antibody and immunoblotted with rabbit FANCA antiserum 89 and anti-flag.

The FANCF region homologous to ROM is not conserved and not essential for function – FANCF was first described as a novel protein with homology to the prokaryotic RNA-binding protein ROM [5]. The sequence alignment shows that the FANCF region homologous to ROM (amino acids 145-209) is not well conserved between human and Xenopus FANCF. Furthermore, the amino acids essential for RNA binding (Gln-166 and Arg-173) are not identical in xFANCF. To further study this region, Arg-173 was changed into alanine and the mutant protein was tested for its ability to complement the MMC hypersensitivity of FA-F cells. Although this amino acid change abolished RNA binding in ROM [37], it did not have an effect on the activity of FANCF (Fig. 3). Similarly, FANCF mutants of the glutamic acid that is identical between hFANCF and ROM (E172A and E172P) were still functional. These data suggest that the FANCF region homologous to ROM is unlikely to have an essential ROM-like function.

The C-terminus of FANCF binds directly to FANCG – Relatively high sequence conservation was found in the C-terminus of FANCF (Fig. 1). To study the function of this conserved region, several deletion constructs were generated and tested for their ability to complement the MMC hypersensitive phenotype of FA-F cells. Surprisingly, FANCF mutants with C-terminal deletions of 23, 31 or 40 amino acids all complemented the MMC hypersensitivity of FA-F cells (Fig. 4A), although the survival curves were slightly shifted towards a higher sensitivity when more amino acids were deleted. Consistent with the notion that FANCD2 monoubiquitination is required for MMC resistance, FA-F cells transfected with the delC31 mutant expressed monoubiquitinated FANCD2 under normal conditions and after MMC treatment (data not shown). Co-immunoprecipitation experiments showed that the FANCF deletion mutants had a reduced interaction with FANCA and FANCG when compared to wild type FANCF (Fig. 4B). Deletion of 31 or 40 amino acids completely abolished the interaction in this assay. Since FANCF has been shown to interact directly with FANCG [15, 38], we tested the interaction between FANCG and the deletion mutants in a yeast two-hybrid assay. In both co-transformation and mating experiments the mutants with a deletion of 31 or 40 amino acids had a strongly reduced FANCG binding (Table 1). These results demonstrate that the last 31-40 amino acids in the C-terminus of FANCF are important for the direct interaction with FANCG. By interfering with the binding to FANCG, the interaction with FANCA is also disturbed, indicating that the C-terminus of FANCF is binding the FANCA/FANCG subcomplex through FANCG.

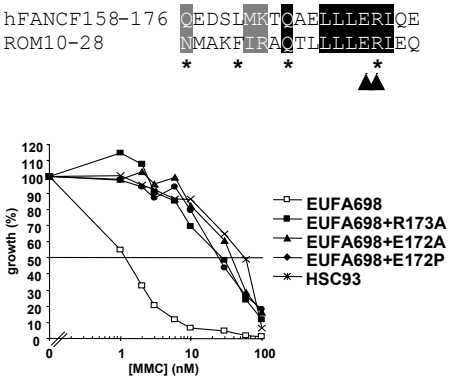


Fig. 3. The FANCF region homologous to ROM is not required to correct the MMC hypersensitivity of FA-F cells**.** Sequence alignment showing the homology between human FANCF and the prokaryotic RNA-binding protein ROM. The asterisks point out the amino acids essential for RNA-binding activity in ROM and the triangles indicate the amino acids that were mutated. The MMC hypersensitive phenotype of the FA-F lymphoblastoid cell line EUFA698 is corrected to wild type levels by stable transfection with cDNAs encoding the FANCF mutants R173A, E172A and E172P. Lymphoblastoid cell line HSC93 is shown as a wild-type control.

Table 1. Yeast two-hybrid data of FANCG and the C-terminal deletion mutants of FANCF

	DNA-BD: pGBK-FANCG DNA-AD: pGAD-FANCF-delC	Reporter gene activation
Co-transformations 1 and 2	FANCF-delC23	+++
	FANCF-delC31	+
	FANCF-delC40	+
Mating assays 1 and 2	FANCF-delC23	+++
	FANCF-delC31	+
	FANCF-delC40	-

BD= GAL4 binding-domain and AD= GAL4 activation-domain

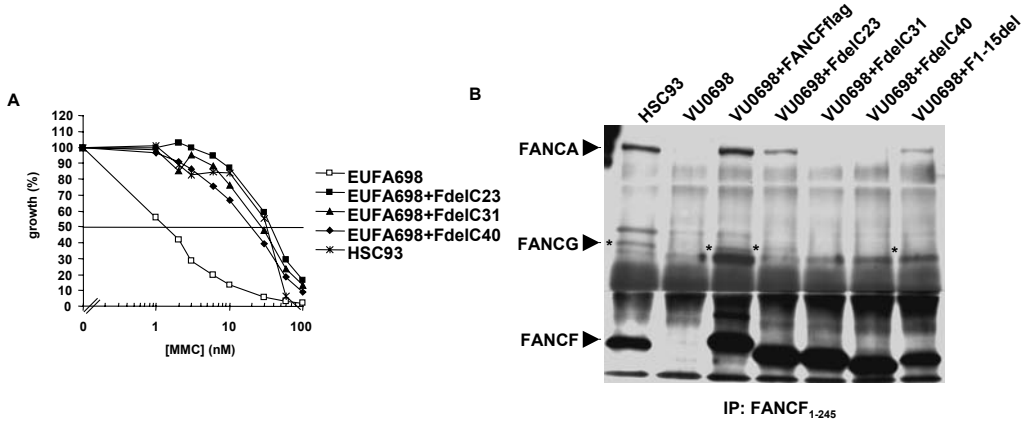


Fig. 4. Deletion mutants of the C-terminus of FANCF are functional, but do not stably interact with FANCA and FANCG. (A) The MMC hypersensitive phenotype of the FA-F cell line EUFA698 is corrected after stable transfection with FANCF mutants that lack 23 (FdelC23), 31 (FdelC31) or 40 (FdelC40) amino acids of the C-terminus. Lymphoblastoid cell line HSC93 is shown as a wild-type control. (B) Cell lysates from wild-type (HSC93), FA-F (EUFA698) and the indicated stably transfected FA-F lymphoblasts were immunoprecipitated with guinea pig anti-$FANCF_{1-245}$ and immunoblotted with rabbit FANCA antiserum 89, rabbit anti-$FANCG_{83-622}$ and rabbit anti-$FANCF_{1-374}$ to show precipitated FANCA, FANCG and FANCF. The asterisks indicate the position of FANCG.

The N-terminus of FANCF is involved in binding of the FANCC/FANCE subcomplex- We next evaluated the function of a deletion mutant that lacked the highly conserved first 15 amino acids of FANCF. The mutant FANCF protein was not able to complement the MMC hypersensitivity of FA-F cells, despite its proper expression and nuclear localization (Fig. 5). Although this mutant protein interacted with FANCA and FANCG, the interaction was weak compared to that of wild type FANCF (Fig. 4B and Fig. 8B).

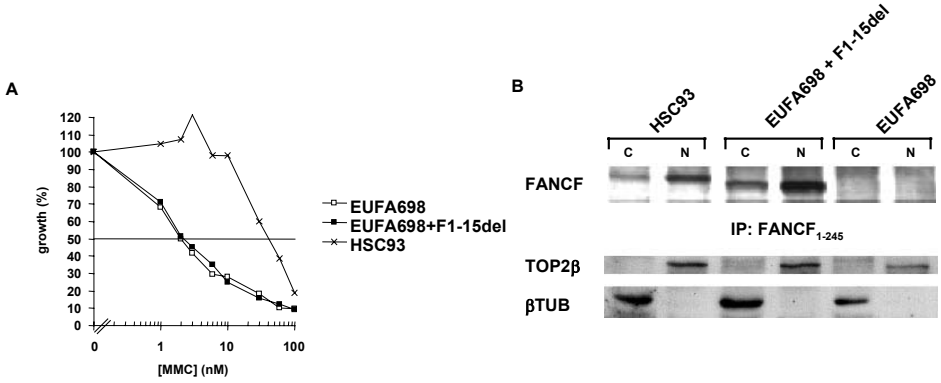


Fig. 5. Deletion of amino acids 1 to 15 inactivates FANCF, without affecting the nuclear localization. (A) The MMC hypersensitive phenotype of the FA-F cell line EUFA698 is not corrected after stable transfection of cDNA encoding a FANCF mutant from which the first N-terminal 15 amino acids were deleted (F1-15del). Lymphoblastoid cell line HSC93 is shown as a wild-type control. (B) Subcellular localization of FANCF mutant F1-15del. Equal amounts of cytoplasmic (C) and nuclear (N) proteins from wild-type (HSC93), FA-F (EUFA698) lymphoblasts and EUFA698 stably transfected with FANCF mutant F1-15del (EUFA698 + F1-15del) were immunoprecipitated with anti-$FANCF_{1-245}$. Precipitated FANCF was visualized by immunoblotting with rabbit anti-$FANCF_{1-374}$. Topoisomerase II β (TOP2β) and β-tubulin (βTUB) antibodies were used as nuclear and cytoplasmic markers, respectively.

Unlike wild-type FANCF, which can be immunoprecipitated in a complex with FANCE, we did not detect any FANCE in a complex with the N-terminal deletion mutant of FANCF using antisera against FANCE (Fig. 6A). Furthermore, the mutant protein did not restore the interaction among FANCE, FANCA and FANCG that is observed in wild-type cells. Similarly, the C-terminal deletion mutant lacking the last 31 amino acids was unable to bind FANCE and could not form a stable FA complex (Fig. 6A). In a reciprocal experiment, FANCC and FANCE did not co-immunoprecipitate with the N-terminal deletion mutant (Fig. 6B). Taken together, these results indicate that the N-terminus of FANCF is essential for its function and is involved in the binding of FANCC and FANCE.

An important question for functional studies is how these proteins associate in the core complex. Previously, no direct interaction between FANCF and FANCC or FANCE has been found in yeast two-hybrid assays [15, 38], suggesting that the association with FANCC and FANCE is indirect. We investigated these interactions in a mammalian two-hybrid assay and also in this assay FANCF did not interact with FANCC or FANCE (Fig. 6C), while an interaction between FANCC and FANCE was observed. We then sought to determine if the N-terminus of FANCF might be involved in the binding of this FANCC/FANCE subcomplex, using a mammalian three-hybrid (M3H) system, with 293 cells stably overexpressing FANCC. In this assay, we observed a strong induction of the reporter gene when FANCF was cotransfected with FANCE (Fig. 6D), indicating that FANCC acts as a molecular bridge between FANCF and FANCE. As a control experiment, we used a 293 cell line stably overexpressing FANCC mutant L554P, which is defective in binding FANCE [16, 38] and found no interaction between FANCF and FANCE (data not shown). To further extend these observations, we tested the FANCF N- and C-terminal deletion mutants in the M3H assay. In agreement with our co-immunoprecipitation experiments, the FANCF mutant 1-15del completely failed to interact with FANCE, whereas the FANCF mutant delC31 had a reduced interaction (Fig. 6D). Collectively, these observations indicate that the first 15 amino acids of FANCF are essential for the direct binding of the FANCC/FANCE subcomplex. Furthermore, they demonstrate that this interaction also partially depends on the C-terminal part of the FANCF protein.

The N-terminus of FANCF contains several residues essential for its function – Several missense mutants were generated to delineate amino acids in the N-terminal region of FANCF essential for its interaction with the other FA proteins and for its ability to complement the MMC defect in FA-F cells (see Table II). Secondary structure prediction programs indicate that the first 15 amino acids of human FANCF form an amphipathic α-helix (Fig. 7), with a strong negative charge at one side of the helix (Glu-2, Asp-9 and Glu-13). In the *Xenopus* homolog a similar structure was predicted, with Glu-6, Asp-9 and Glu-13 at the negative side of the helix.

These negatively charged amino acids were either changed into alanines (E2A+D9A+E13A) or replaced by positively charged lysines (E2K+D9K+E13K) and the mutants were tested for their ability to complement the MMC hypersensitivity of FA-F cells, and for their ability to interact with FANCA. Both mutants were able to complement the MMC defect (Fig. 8A), but like the N-terminal deletion mutant, the E2K+D9K+E13K mutant appeared to have a reduced interaction with FANCA (Fig. 8B). Since the C-terminus of FANCF is also involved in the interaction with FANCA (Fig. 4B), we tested mutants with a combination of mutation and deletions to determine if the N- and C-terminal conserved regions of FANCF act in concert. When the last 23 C-terminal amino acids were removed, the E2A+D9A+E13A mutant was still biologically active, whereas the E2K+D9K+E13K mutant became inactive (Fig. 8C). We then removed 31 C-terminal amino acids of these mutants and found that this completely abolished the activity of both mutants (Fig. 8D). Thus, reversing the charges of the residues in the FANCF α-helix is more disruptive than neutralizing the negatively charged residues. We infer that the three negatively charged amino acids in the first N-terminal α-helix of FANCF are functionally important and that FANCF has partially interdependent functional terminal domains.

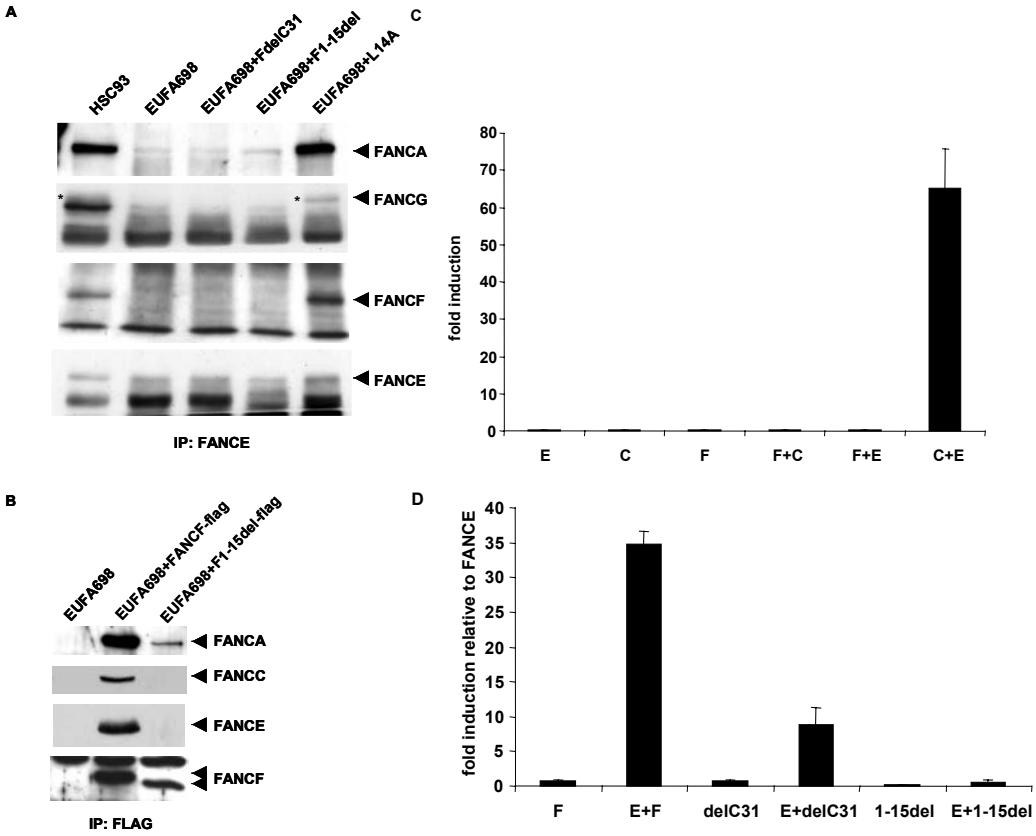


Fig. 6. Interaction of FANCF mutants with the FANCC/FANCE subcomplex. (A) Cell lysates from wild-type (HSC93), FA-F (EUFA698) lymphoblasts and FA-F cells stably transfected with the indicated FANCF mutants were immunoprecipitated with anti-FANCE and immunoblotted with anti-FANCA serum 89, anti-$FANCG_{83\text{-}622}$, anti-$FANCF_{1\text{-}374}$ and anti-FANCE to visualize FANCA, FANCG, FANCF and FANCE respectively. The asterisks indicate the position of FANCG. (B) EUFA698 cells were stably transfected with cDNAs encoding the FANCF mutant F1-15del-flag and wild-type FANCF-flag in expression vector pIRESneo. Immunoprecipitation was performed with anti-flag resin and immunoprecipitated FA proteins were visualized by immunoblotting with anti-FANCA serum 89, anti-$FANCC_{106\text{-}558}$, anti-FANCE and anti-$FANCF_{1\text{-}374}$. (C) Mammalian two-hybrid assay indicating a lack of direct interaction between FANCF and FANCC or FANCE. Fold induction is expressed relative to the luciferase activity obtained with empty vectors (pM and pVP16). (D) Mammalian three-hybrid assay illustrating the interaction of FANCF and FANCF mutants with the FANCC/FANCE subcomplex. Fold induction is expressed relative to the luciferase activity obtained with the pVP16-AD-FANCE vector alone. Results shown are mean ± standard deviation (SD) and derived from an experiment in triplicate.

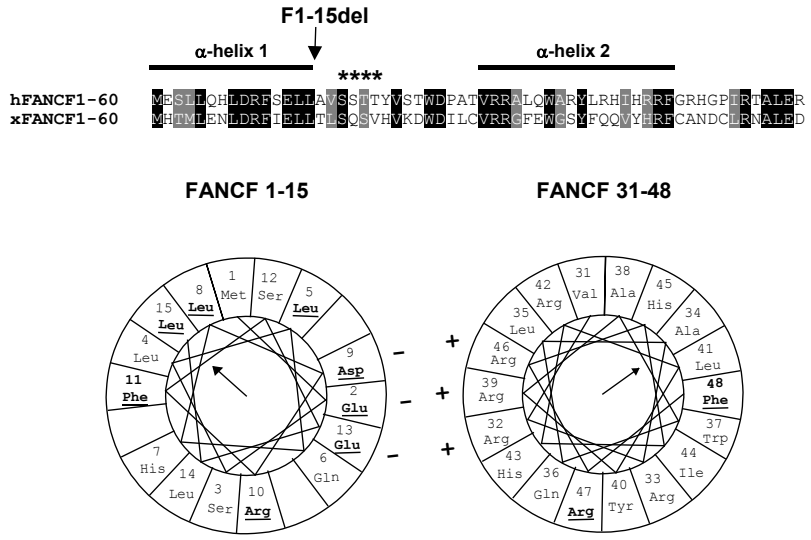


Fig. 7. The N-terminus of FANCF contains two amphipathic α-helices. Sequence alignment showing the homology between the N-terminus of human and *Xenopus laevis* FANCF. Two amphipathic α-helices were predicted by Jpred and DNAsis software. The arrows indicate the hydrophobic moment of the helices. The functionally important amino acids are in bold and underlined. Asterisks (*) indicate the serine-threonine stretch.

The hydrophobic side of the N-terminal α-helix contains three conserved leucines (Leu-5, Leu-8 and Leu-15). When these leucines were replaced by alanines, the mutant protein was functional in an MMC test (Fig. 9A). However, this mutant was totally inactivated upon removal of the 31 C-terminal amino acids (Fig. 9A), hinting that the conservation at the hydrophobic side of the helix is of functional significance. We then examined the ability of the FANCF mutant L5A+L8A+L15A (with an intact C-terminus) to interact with the FANCC/FANCE subcomplex in the M3H assay and found that this mutant failed to activate transcription of the luciferase reporter gene (Fig. 9C). These results suggest that the three leucines residues are involved in the direct binding of the FANCC/FANCE subcomplex.

We further investigated N-terminal mutants, in which combinations of the conserved arginine (Arg-10) or phenylalanine (Phe-11) residues were modified. All the single mutants were able to complement the MMC hypersensitivity of FA-F cells (Table II) and displayed a normal interaction with FANCA (Fig. 8B). These mutants were still functional when the C-terminal 31 amino acids were removed (Table II). However, a double mutant in which both Arg-10 and Phe-11 were changed into alanine (R10A+F11A) had a reduced biological activity, but only in combination with a C-terminal deletion of 31 amino acids (Fig. 9B). A possible explanation for the partial inactivation of the protein is the prediction of a similar α-helix (amino acids

31-48), in which a conserved arginine (Arg-47) and phenylalanine (Phe-48) are found (Fig. 7). These amino acids might be able to partially compensate for the amino acid changes in the first α-helix. To test this hypothesis, we constructed a mutant in which the arginines and phenylalanines in both helices were replaced by alanines (R10A+F11A+R47A+F48A). When combined with a deletion of the C-terminus, this mutant was totally inactive (Fig. 9B). A mutant in which only Arg-47 and Phe-48 were changed (R47A+F48A) was functional (Fig. 9B). These results indicate that the conserved arginines (Arg-10 and Arg-47) and phenylalanines (Phe-11 and Phe-48) in the first and second α-helix may form a functional unit. The M3H assay was used to show that the R10A+F11A+R47A+F48A mutant (with an intact C-terminus) was unable to interact with the FANCC/FANCE subcomplex (Fig. 9C), suggesting that these arginine and phenylalanine residues participate in this direct interaction.

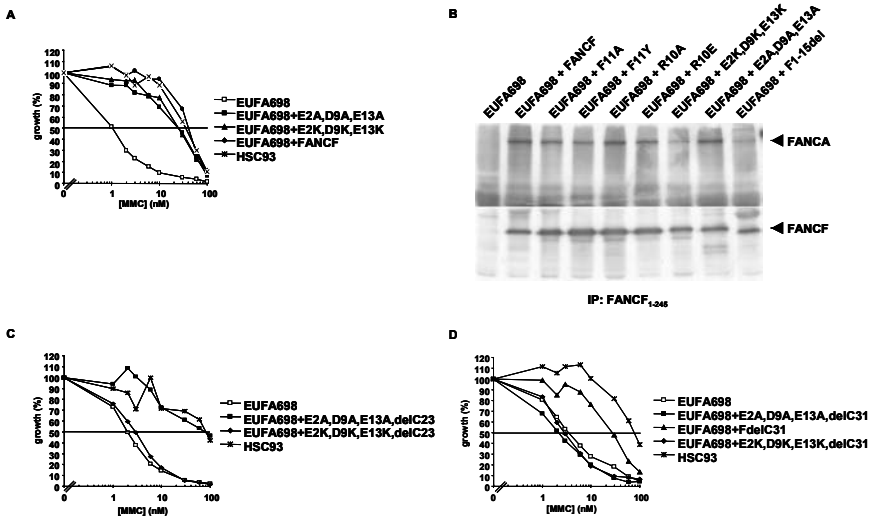


Fig. 8. The negatively charged amino acids in the first N-terminal α-helix of FANCF have an important function. (A) The MMC hypersensitive phenotype of the FA-F cell line EUFA698 is corrected after stable transfection of cDNAs encoding FANCF mutants in which the negatively charged amino acids Glu-2, Asp-9 and Glu-13 were mutated. Lymphoblastoid cell line HSC93 is shown as a wild-type control. (B) Cell lysates from FA-F lymphoblasts (EUFA698) and FA-F lymphoblasts stably transfected with the indicated FANCF mutants were immunoprecipitated with guinea pig anti-$FANCF_{1-245}$ and immunoblotted with rabbit FANCA antiserum 89 and rabbit anti-$FANCF_{1-374}$ to show the interaction between the different FANCF mutants and FANCA. (C) Deletion of 23 amino acids from the FANCF mutants in which the negatively charged amino acids were mutated does only affect the ability of the E2K+D9K+E13K mutant to complement the MMC hypersensitive phenotype of the FA-F cell line EUFA698. (D) The FANCF mutants in which the negatively charged amino acids were mutated are unable to complement the MMC hypersensitive phenotype of the FA-F cell line EUFA698 upon removal of the C-terminal 31 amino acids.

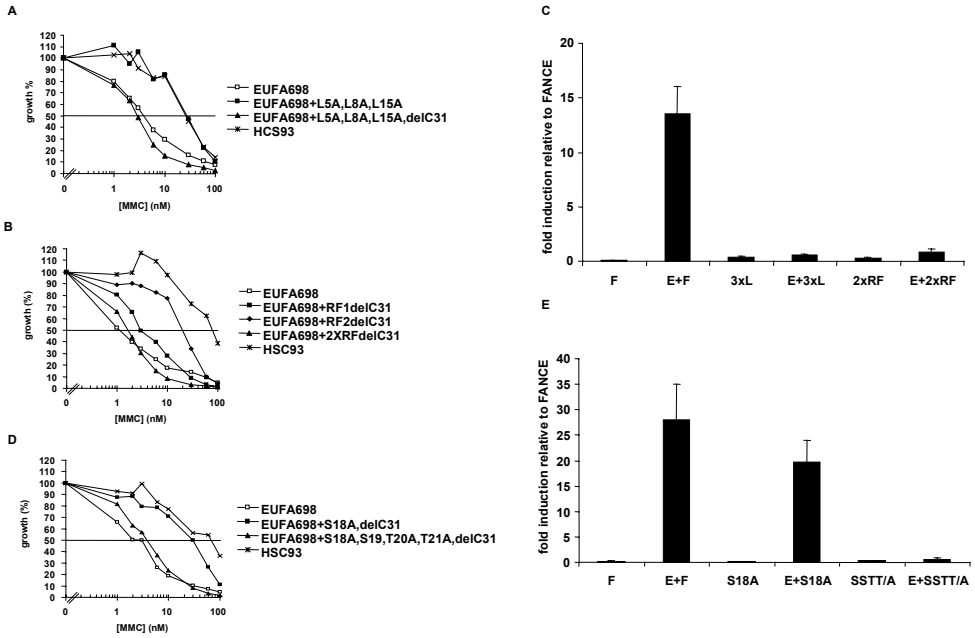


Fig. 9. Other functional important amino acids in the N-terminus of FANCF. (A) A FANCF mutant in which Leu-5, Leu-8 and Leu-15 are changed into alanines is not able to complement the MMC hypersensitive phenotype of EUFA698 cells after removal of the C-terminal 31 amino acids. (B) Effect of arginine and phenylalanine-substitutions in FANCF on the ability to complement the MMC hypersensitive phenotype of EUFA698 cells. FANCF mutants R10A+F11A and R10A+F11A+R47A+R48A are inactivated upon removal of the C-terminus**.** (C) Mammalian three-hybrid assay illustrating the disturbed interaction between the FANCF mutants 3xL (L5A+L8A+L15A) and 2xRF (R10A+F11A+R47A+R48A) and the FANCC/FANCE subcomplex. Results shown are mean ± standard deviation (SD) and derived from an experiment in triplicate. Fold induction is expressed relative to the luciferase activity obtained with the pVP16-AD-FANCE vector alone. (D) A FANCF mutant in which Ser-18, Ser19, Thr-20 and Thr-21 are changed into alanine is not able to complement the MMC hypersensitive phenotype of EUFA698 cells after removal of the C-terminal 31 amino acids. (E) Mammalian three-hybrid assay showing interaction of the FANCC/FANCE subcomplex with wild type FANCF and FANCF mutant S18A, which is lost in the FANCF mutant S18A-S19A-T20A-T21A (SSTT/A).

In order to investigate whether the two predicted N-terminal α-helices are really being formed and function as a unit, several leucine to proline mutants were made to disrupt the helical structure of this region (Table II). The disruption of the first α-helix by substitution of Leu-8 or Leu-14 as well as the disruption of the second α-helix by substitution of Leu-41 interfered with the biological activity of the protein, but again only when the C-terminal 31 amino acids were removed. As a control, we replaced Leu-14 by alanine and deleted its C-terminus, and found that this did not affect the function of the protein. These results imply that the two N-terminal α-helices in FANCF form a functionally important structure, which cooperates with the C-terminus of the protein.

A serine and threonine stretch (Ser-18, Ser-19, Thr-20, Thr-21) is located just after the first N-terminal α-helix. To investigate the functional significance of this region we changed these amino acids into alanines. This mutant was again inactive after

removal of its C-terminus (Fig. 9D). However, a mutant (S18A) in which only the conserved Ser-18 was changed to an alanine appeared to be functional (Fig. 9D). Subsequent analysis of these FANCF mutants (with an intact C-terminus) in the M3H assay showed that only the mutant S18A was able to activate the luciferase reporter gene (Fig. 9E). These data suggest that the serine and threonine stretch is of functional importance and is involved in the binding of the FANCC/FANCE subcomplex.

Amino acids 76 to 84 are not essential for the function of FANCF – FA-F patient EUFA698 has a 23 bp deletion in the FANCF gene resulting in a frame shift at codon 77 (5). Since the deleted region is poorly conserved between human and Xenopus FANCF, we tried to overcome the defect by selection for a phenotypic revertant *in vitro*. After long-term culture of a lymphoblastoid cell line from patient EUFA698 in the presence of 15 nM MMC a MMC resistant cell line was obtained. Sequence analysis indicated that a deletion of nucleotide 225 (225delT) had restored the FANCF reading frame in one allele of the reverted cell line (Fig. 10). In the mutant protein, Pro-76 is changed into a glutamine while amino acids 77 to 84 are absent. This FANCF mutant is expressed and has a normal interaction with FANCA (Fig. 10). These data therefore indicate that amino acids 76 to 84 are not essential for the complementing activity of the FANCF protein and support the idea that the structure of FANCF is rather flexible.

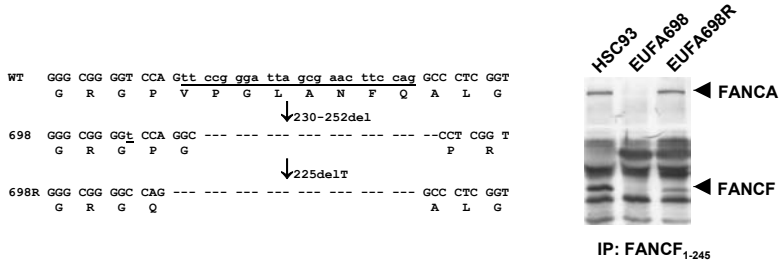


Fig. 10. Amino acids 76 to 84 are not essential for the function of FANCF. A 23 base pair deletion (230-252del) in FA-F patient EUFA698 results in a frame shift at codon 77 (Val-77). The FANCF reading frame is restored by a deletion of nucleotide 225 in revertant cell line EUFA698R, which changes Pro-76 into Gln and deletes amino acids 77 to 84. Cell lysates from wild type lymphoblasts (HSC93), FA-F lymphoblasts (EUFA698) and the reverted FA-F lymphoblasts (EUFA698R) were immunoprecipitated with guinea pig anti-$FANCF_{1-245}$ and immunoblotted with rabbit FANCA antiserum 89 and rabbit anti-$FANCF_{1-374}$ to show the interaction between FANCF and FANCA.

Table 2. Activity of the FANCF mutants in different assays

FANCF mutant	MMC test	FANCA interaction (IP)	MMC test upon removal of the C-terminus	FANCC/FANCE interaction (M3H)
delC31	+	-	na	↓
1-15del	-	↓	nd	-
Q6A	+	+	+	nd
D9A + R10A	+	nd	+	nd
R10A	+	+	+	nd
R10E	+	+	+	nd
F11A	+	+	+	nd
F11Y	+	+	+	nd
R10A + F11A	+	+	↓	-
F11A + E13A	+	+	↓	nd
R47A + F48A	+	+	+	nd
R10A + F11A + R47A + F48A	+	nd	-	-
E2A + D9A + E13A	+	+	-	nd
E2K + D9K + E13K	+	↓	-	nd
L5A + L8A + L15A	+	↓	-	-
L8P	+	nd	-	nd
L14P	+	nd	-	nd
L14A	+	+	+	nd
L41P	+	+	-	nd
L8P + L41P	+	nd	nd	nd
L14P + L41P	+	nd	nd	nd
S18A	+	+	+	+
S18A + S19A + T20A + T21A	+	+	-	-

+ = normal activity; ↓ = reduced activity; - = no activity; na= not applicable; nd = not determined

Discussion

Highly conserved amino acids are generally thought to be critical for a protein's function. By using an extensive site-directed mutagenesis approach we have investigated the functional importance of amino acids that are conserved between human and *Xenopus laevis* FANCF. Our results demonstrate that the C-terminus of FANCF binds directly to FANCG and is required for stable interaction with FANCA, FANCC and FANCE. Furthermore, we found that the N-terminus of FANCF is essential for the direct interaction with the FANCC/FANCE subcomplex, and is important for stabilizing the complex formation with FANCA and FANCG. This study shows that FANCF is an adaptor protein, which keeps the other components of the FA core complex together in such a way that they can perform their function.

A direct interaction between FANCF and FANCG has been demonstrated using the yeast two-hybrid system [15] and the carboxy terminus of FANCF (amino acids 243-374) was shown to be essential for this interaction [38]. In the current study, we found that the interaction site is specifically located in the strongly conserved last 31 amino acids (amino acids 343-374) of FANCF and that this particular region provides stability to the FA core complex. Surprisingly, the C-terminal deletion mutants were still able to complement the MMC hypersensitivity of FA-F cells and to support the monoubiquitination of FANCD2. The ability of the FANCFdelC31 mutant to complement FA-F cells was not a result of overexpression of this protein, since low expression levels of this mutant protein by a leaky inducible vector (pMEP4) or by stable integration (pIRESneo) still corrected the MMC hypersensitivity [J.P. de Winter et al., unpublished data]. Since the mutant FANCFdelC31 also corrects the MMC hypersensitivity of FA-F patient cell line EUFA1228 [J.P. de Winter et al., unpublished data], the complementing activity of this mutant is also not cell line dependent. We infer from our results that the FANCFdelC31 mutant protein has a reduced affinity for the other members of the complex, undetectable in co-immunoprecipitation experiments because of the stringent conditions of this assay, but sufficient for a functional FA pathway. This conclusion is supported by our data showing an interaction between the FANCFdelC31 mutant and the FANCC/FANCE subcomplex in the mammalian three hybrid assay, which was undetectable in co-immunoprecipitation experiments.

The first 15 amino acids of FANCF are predicted to form an α-helix sharing 60% homology with *Xenopus* FANCF. This region is essential for the function of FANCF, since a FANCF mutant in which these amino acids were deleted failed to correct the MMC hypersensitivity of EUFA698 cells. This FANCF mutant appeared to have a normal nuclear localization and a reduced interaction with FANCA and FANCG, indicating that this particular region in FANCF is not directly involved in the nuclear

targeting of FANCF, but crucial for the stable binding to FANCA and FANCG. Interestingly, the FANCF mutant 1-15del failed to co-precipitate with FANCC and FANCE, suggesting that this particular N-terminal region in FANCF is required for the presence of FANCC and FANCE within the FA core complex. The mammalian three-hybrid system confirms this finding and provides strong evidence for direct interaction between the first 15 amino acids of FANCF and the FANCC/FANCE subcomplex.

To identify functionally relevant residues in the N-terminus of FANCF, we generated several mutations in the first 48 amino acids that contain two predicted α-helices. Surprisingly, we could not obtain inactive or partially functional mutants without the additional removal of the last 31 amino acids when we assayed for the ability to complement the MMC hypersensitive phenotype of FA-F cells. This implies that the FANCA-FANCG interaction needs to be disturbed before amino acid substitutions in the N-terminus of FANCF show an effect in this assay. However, when tested in the M3H assay, mutations in the N-terminus of FANCF were sufficient to interfere with the binding of the FANCC/FANCE subcomplex. Apparently, the binding of FANCA and FANCG to the C-terminus of FANCF stabilizes the interaction of factors that bind to the N-terminus of FANCF, in particular the FANCC/FANCE subcomplex. By removing the C-terminus of the FANCF protein we interfered with this stabilizing effect. In the mammalian three hybrid assay we do not need to interfere with the binding of the FANCA/FANCG subcomplex, since in this assay FANCA and FANCG are not present. What we have essentially tested in the mammalian three hybrid assay and the MMC test with the C-terminally truncated N-terminal mutants is the same; the interaction with the FANCC/FANCE subcomplex in the absence of the FANCA/FANCG subcomplex. Accordingly, the results of both assays are the same. The detection of a weak interaction between FANCA and FANCE [Ref 16 and Fig. 6A] and between FANCA and FANCC [14] in co-immunoprecipitation experiments on lysates of FA-F patients indicates that these FA proteins have some affinity for each other, but that these interactions need to be stabilized by the presence of FANCF. The results of our experiments suggest that binding of the FANCC/FANCE subcomplex to the N-terminus of FANCF (in case of the C-terminal deletion mutants) or binding of the FANCA/FANCG subcomplex to C-terminus of FANCF (in case of the N-terminal missense mutants) is sufficient to activate the FA pathway.

From the non-functional mutants we identified four critical regions defined by a group of residues essential for the proper function of FANCF. One region consists of 3 consecutive negatively charged residues (Glu-2, Asp-9, Glu-13) located at one side of the first α-helix. It is tempting to speculate that this region interacts through salt

bridges with the 3 positively charged residues (Arg-32, Arg-39, Arg-46) located at one side of the second α-helix to form a functional domain. Another important region involves 3 conserved hydrophobic amino acids on one side of the first α-helix (Leu-5, Leu-8, Leu-15), which appear to form a binding site for the FANCC/FANCE subcomplex. The conserved arginines (Arg-10 and Arg-47) and phenylalanines (Phe-11 and Phe-48) in the first and second α-helix form the third critical region in the N-terminus of FANCF. These 4 residues combined are also involved in the association with the FANCC/FANCE subcomplex and may either be a part of the interaction domain or stabilize this domain. A fourth critical region in the FANCF protein is formed by a stretch of serine-threonine residues in the loop region between the two α-helices. Although there is no evidence as yet for posttranslational modification of FANCF, this specific region could be the target for phosphorylation by a serine/threonine kinase. Alternatively, these serines and threonines might be involved in protein-protein interaction through the formation of hydrogen bonds. Our data show that this particular region plays a role in the direct interaction with the FANCC/FANCE subcomplex. Finally, the FANCF region homologous to the prokaryotic RNA binding protein ROM (amino acid residues 145-209) does not seem to have an essential ROM-like function. The FANCF mutants R172A, E172P and R173A (equivalent in ROM to amino acids Q24 and R25, respectively) did not affect the function of FANCF.

We realize that these mutagenesis studies need to be interpreted with caution, because of the risk to affect the general structure of the protein. However, since amino acid changes in the N-terminus and even deletions of the C-terminus of FANCF were tolerated, the FANCF structure seems to be very stable and flexible. Also the fact that a deletion of the C-terminus only inactivates a subset of the N-terminal FANCF mutants indicates that it is very difficult to affect the structure of the FANCF protein. Consistent with this idea is that none of the identified FA-F patients have missense mutations in FANCF, but rather have nonsense mutations or deletions [5]. Furthermore, an *in vitro* revertant of EUFA698 cells expresses a mutant FANCF protein that lacks amino acids 76 to 84, which is still functional. The only mutation that might have changed the structure of the FANCF protein is the N-terminal deletion mutant, since this is the only single mutant that is unable to restore the MMC hypersensitivity of FA-F cells.

Our results suggest a sequential recruitment of the FA gene products in which the FANCF protein has a key role by linking the different subcomplexes (FANCA/FANCG and FANCC/FANCE) and possibly other components together. Therefore, we propose that FANCF acts as an adaptor protein that plays a key role in

the proper assembly of the FA core complex. We infer from our and published data that FANCF stabilizes the interaction between the FANCC/FANCE subcomplex and the FANCA/FANCG subcomplex and locks the whole FA core complex in a conformation that is essential to perform its function.

Acknowledgements
We thank Bianca Barrett and Roderik van Beers for technical assistance. This study was supported by the Fanconi Anemia Research Fund (Eugene, OR), the National Institutes of Health, the Medical Research Council (UK), the Dutch Cancer Society, and the Netherlands Organization for Health Research and Development.

References

1 Auerbach AD, Buchwald M, Joenje H (2002) *Genetics of Human Cancer*, Vogelstein B, Kinzler KW, Ed., pp.317-322, McGraw-Hill, NY

2 Joenje H, Patel KJ (2001). The emerging genetic and molecular basis of Fanconi anaemia. *Nat Rev Genet* **2:** 446–457.

3 Levitus M, Rooimans MA, Steltenpool J, Cool NF, Oostra AB, Mathew CG, Hoatlin ME, Waisfisz Q, Arwert F, de Winter JP, Joenje H (2004). Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes. *Blood* **103:** 2498–2503.

4 de Winter JP, Waisfisz Q, Rooimans MA, van Berkel CG, Bosnoyan-Collins L, Alon N, Carreau M, Bender O, Demuth I, Schindler D, Pronk JC, Arwert F, Hoehn H, Digweed M, Buchwald M, Joenje H (1998). The Fanconi anaemia group G gene FANCG is identical with XRCC9. *Nat Genet* **20:** 281–283.

5 de Winter JP, Rooimans MA, van der Weel L, van Berkel CG, Alon N, Bosnoyan-Collins L, de Groot J, Zhi Y, Waisfisz Q, Pronk JC, Arwert F, Mathew CG, Scheper RK, Hoatlin ME, Buchwald M, Joenje, H (2000). The Fanconi anaemia gene FANCF encodes a novel protein with homology to ROM. *Nat Genet* **24:** 15–16.

6 de Winter JP, Léveillé F, van Berkel CG, Rooimans MA, van der Weel L, Steltenpool J, Demuth I, Morgan NV, Alon N, Bosnoyan-Collins L, Lightfoot J, Leegwater PA, Waisfisz Q, Komatsu K, Arwert F, Pronk JC, Mathew CG, Digweed M, Buchwald M, Joenje, H (2000). Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am J Hum Genet* **67:** 1306–1308.

7 The Fanconi Anaemia/Breast Cancer Consortium (1996). Positional cloning of the Fanconi anaemia group A gene. *Nat Genet* **14:** 324–328.

8 Howlett NG, Taniguchi T, OlsonS, Cox B, Waisfisz Q, de Die-Smulders C, Persky N, Grompe M, Joenje H, Pals G, Ikada H, Fox EA, D'Andrea AD (2002). Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297:** 606–609.

9 Lo Ten Foe JR, Rooimans MA, Bosnoyan-Collins L, Alon N, Wijker M, Parker L, Lightfoot J, Carreau M, Callen DF, Savoia A, Cheng NC, van Berkel CG, Strunk MH, Gille JJ, Pals G, Kruyt FA, Pronk JC, Arwert F, Buchwald M, Joenje H (1996). Expression cloning of a cDNA for the major Fanconi anaemia gene, FAA. *Nat Genet* **14:** 320–323.

10 Meetei AR, de Winter JP, Medhurst AL, Wallisch M, Waisfisz Q, van de Vrugt HJ, Oostra AB, Yan Z, Ling C, Bishop CE, Hoatlin ME, Joenje H, Wang W (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35:** 165–170.

11 Strathdee CA, Gavish H, Shannon WR, Buchwald M (1992). Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **358:** 434.

12 Timmers C, Taniguchi T, Hejna J, Reifsteck C, Lucas L, Bruun D, Thayer M, Cox B, Olson S, D'Andrea AD, Moses R, Grompe M. (2001). Positional cloning of a novel Fanconi anemia gene, FANCD2. *Mol Cell* **7:** 241–248.
13 Blom E, van de Vrugt HJ, Vries Y, de Winter JP, Arwert F, Joenje H (2004). Multiple TPR motifs characterize the Fanconi anemia FANCG protein. *DNA Repair* **3:** 77–84.
14 de Winter JP, van der Weel L, de Groot J, Stone S, Waisfisz Q, Arwert F, Scheper RJ, Kruyt FA, Hoatlin ME, Joenje H (2000). The Fanconi anemia protein FANCF forms a nuclear complex with FANCA, FANCC and FANCG. *Hum Mol Genet* **9:** 2665–2674.
15 Medhurst AL, Huber PA, Waisfisz Q, de Winter JP, Mathew CG (2001). Direct interactions of the five known Fanconi anaemia proteins suggest a common functional pathway. *Hum Mol Genet* **10:** 423–429.
16 Pace P, Johnson M, Tan WM, Mosedale G, Sng C, Hoatlin M, de Winter J, Joenje H, Gergely F, Patel KJ (2002). FANCE: the link between Fanconi anaemia complex assembly and activity. *EMBO J* **21:** 3414–3423.
17 Waisfisz Q, de Winter JP, Kruyt FA, de Groot J, van der Weel L, Dijkmans LM, Zhi Y, Arwert F, Scheper RJ, Youssoufian H, Hoatlin ME, Joenje H (1999). A physical complex of the Fanconi anemia proteins FANCG/XRCC9 and FANCA. *Proc Natl Acad Sci USA* **96:** 10320–10325.
18 Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J, Grompe M, D'Andrea AD (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7:** 249–262.
19 Taniguchi T, Garcia-Higuera I, Andreassen PR, Gregory RC, Grompe M, D'Andrea AD (2002). S-phase-specific interaction of the Fanconi anemia protein, FANCD2, with BRCA1 and RAD51. *Blood* **100:** 2414–2420.
20 Hussain S, Wilson JB, Medhurst AL, Hejna J, Witt E, Ananth S, Davies A, Masson J Y, Moses R, West SC, de Winter JP, Ashworth A, Jones NJ, Mathew CG (2004). Direct interaction of FANCD2 with BRCA2 in DNA damage response pathways. *Hum Mol Genet* **13:** 1241–1248.
21 Venkitaraman AR (2002). Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108:** 171–182.
22 Yang H, Jeffrey PD, Miller J, Kinnucan E, Sun Y, Thoma NH, Zheng N, Chen PL, Lee WH, Pavletich NP (2002). BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science* **297:** 1837–1848.
23 Taniguchi T, Garcia-Higuera I, Andreassen PR, Gregory RC, Kim ST, lane WS, Kastan MB, D'Andrea AD (2002). Convergence of the fanconi anemia and ataxia telangiectasia signaling pathways. *Cell* **109:** 459–472.

24 Nakanishi K, Taniguchi T, Ranganathan V, New HH, Moreau LA, Stotsky M, Mathew CG, Kastan MB, Weaver DT, D'Andrea AD (2002). Interaction of FANCD2 and NBS1 in the DNA damage response. *Nat Cell Biol* **4:** 913–920.
25 Pichierri P, Averbeck D, Rosselli F (2002). DNA cross-link-dependent RAD50/MRE11/NBS1 subnuclear assembly requires the Fanconi anemia C protein. *Hum Mol Genet* **11:** 2531–2546.
26 D'Amours D, Jackson SP (2002). The Mre11 complex: at the crossroads of dna repair and checkpoint signalling. *Nat Rev Mol Cell Biol* **3:** 317–327.
27 Taniguchi T, Tischkowitz M, Ameziane N, Hodgson SV, Mathew CG, Joenje H, Mok SC, D'Andrea AD (2003). Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors. *Nat Med* **9:** 568–574.
28 Marsit CJ, Liu M, Nelson HH, Posner M, Suzuki M, Kelsey KT (2004). Inactivation of the Fanconi anemia/BRCA pathway in lung and oral cancers: implications for treatment and survival. *Oncogene* **23:** 1000–1004.
29 Narayan G, Arias-Pulido H, Nandula SV, Basso K, Sugirtharaj DD Vargas H, Mansukhani M, Villella J, Meyer L, Schneider A, Gissmann L, Durst M, Pothuri B, Murty VV (2004). Promoter hypermethylation of FANCF: disruption of Fanconi Anemia-BRCA pathway in cervical cancer. *Cancer Res* **64:** 2994–2997.
30 Kupfer GM, Naf D, Garcia-Higuera I, Wasik J, Cheng A, Yamashita T, Tipping A, Morgan N, Mathew CG, D'Andrea AD (1999). A patient-derived mutant form of the Fanconi anemia protein, FANCA, is defective in nuclear accumulation. *Exp Hematol* **27:** 587–593.
31 Kuang Y, Garcia-Higuera I, Moran A, Mondoux M, Digweed M, D'Andrea AD (2000). Carboxy terminal region of the Fanconi anemia protein, FANCG/XRCC9, is required for functional activity. *Blood* **96:** 1625–1632.
32 Nakanishi K, Moran A, Hays T, Kuang Y, Fox E, Garneau D, Monte de Oca R, Grompe M, D'Andrea AD (2001). Functional analysis of patient-derived mutations in the Fanconi anemia gene, FANCG/XRCC9. *Exp Hematol* **29:** 842–849.
33 Pang Q, Christianson TA, Keeble W, Diaz J, Faulkner GR, Reifsteck C, Olson S, Bagby GC (2001). The Fanconi anemia complementation group C gene product: structural evidence of multifunctionality. *Blood* **98:** 1392–1401.
34 Adachi D, Oda T, Yagasaki H, Nakasato K, Taniguchi T, D'andrea AD, Asano S, Yamashita T (2002). Heterogeneous activation of the Fanconi anemia pathway by patient-derived FANCA mutants. *Hum Mol Genet* **11:** 3125–3134.
35 Ishida R, Buchwald M (1982). Susceptibility of Fanconi's anemia lymphoblasts to DNA-cross-linking and alkylating agents. *Cancer Res* **42:** 4000–4006.

36 Joenje H, Lo Ten Foe JR, Oostra AB, van Berkel CG, Rooimans MA, Schroeder-Kurth T, Wegner RD, Gille JJ, Buchwald M, Arwert F (1995). Classification of Fanconi anemia patients by complementation analysis: evidence for a fifth genetic subtype. *Blood* **86:** 2156–2160.

37 Predki PF, Nayak LM, Gottlieb MB, Regan L (1995). Dissecting RNA-protein interactions: RNA-RNA recognition by Rop. *Cell* **80:** 41–50.

38 Gordon SM, Buchwald M (2003). Fanconi anemia protein complex: mapping protein interactions in the yeast 2- and 3-hybrid systems. *Blood* **102:** 136–141.

CHAPTER

5

Bioinformatic analysis of the Fanconi anemia gene products

Background – The bone marrow failure and DNA repair disorder Fanconi anemia (FA) is genetically highly heterogeneous, as defects in each of at least 9 different genes can precipitate this disease: *FANCA, -B, -C, -D1, -D2, -E, -F, -G,* and *-L*. Functional characterization of the individual gene products has been problematic, because most are known as 'orphans', so that little information can be deduced from direct comparison with proteins in the databases. We previously used a bioinformatic approach, based on comparisons with distant orthologs in the bony fishes *Danio rerio* and *Oryzias latipes,* to show the existence of TPR motifs in the FANCG protein. We now extend this analysis to include other FANC proteins and additional software such as coiled-coil prediction and protein fold recognition. FANCD1/BRCA2 was excluded, since this protein has been studied extensively by others.

Results – Distinct features were discovered in four FANC proteins, i.e. FANCL, FANCD2, FANCA, and FANCC. Contrary to what has been reported previously, FANCL contains a RING finger domain rather than a PHD-type zinc finger domain. In addition, FANCD2 may have a coiled coil structure and both FANCA and FANCC are likely composed of multiple HEAT repeat-like motifs.

Conclusions – The bioinformatic strategy as outlined in this paper was successful in extracting information on the domain structure of four FANC proteins. The information allows the postulation of new testable hypotheses on the function of these proteins.

Eric Blom, Fré Arwert & Hans Joenje

Unpublished results

Introduction

Fanconi anemia (FA), a chromosomal instability syndrome characterized by aplastic anemia, developmental defects and cancer predisposition, is genetically a highly heterogeneous disease. To date, 11 complementation groups have been distinguished (FA-A, B, C, D1, D2, E, F, G, I, J, and L), for 9 of which the corresponding genes have been identified (*FANCA*, *-B*, *-C*, *-D1*, *-D2*, *-E*, *-F*, *-G*, and *-L*) [1-10]. Biochemical studies have revealed that most of the FANC proteins assemble in a multisubunit protein 'core complex' composed of FANCA, -B, -C, -E, -F, -G, and –L, which is required for activation via monoubiquitylation of the downstream functioning FANCD2 protein [11, 12]. Interestingly, the breast cancer susceptibility gene products BRCA1 and BRCA2 are also involved in the FA pathway, as the FANCD2 protein forms nuclear repair foci with BRCA1 upon its monoubiquitylation, while *BRCA2* turned out be the same gene as *FANCD1* [11, 13]. Furthermore, at the cellular level, FA cells are extremely sensitive to agents that induce DNA interstrand cross-links, such as mitomycin C and diepoxybutane, suggesting a role for the FA pathway in the handling of this type of DNA lesion.

Despite a considerable progress in FA research, the exact function of the individual FANC proteins is poorly understood. This is partially caused by the fact that most FANC proteins are 'orphans' having no homology to each other nor to any other protein, making a bioinformatic analysis difficult or even impossible. A positive exception has been the FANCG protein, which has recently been characterized as a member of the tetratricopeptide repeat (TPR) family predicting a role in the assembly of the FA protein core complex [14]. Results depended on the application of an in-depth bioinformatic approach, which is used in this report to analyze the remaining FANC proteins. The FANCD1/BRCA2 protein was excluded from this analysis, since its domain structure has already been studied in great detail by others: this protein contains 8 copies of a 39 amino acid repeat, the BRC repeat, which is critical for binding to the DNA recombinational repair protein RAD51 [15, 16].

Methods

Bioinformatic analysis was performed as outlined in the results and discussion section. Database searches as BLAST, PSI-BLAST and the rpsblast tool were run at the website of the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/BLAST/). For the prediction of putative coiled coil regions the COILS program at http://www.ch.embnet.org/software/COILS_form.html was used and for the prediction of protein families the Pfam server at http://pfam.wustl.edu/. For protein fold recognition using mGenThreader the server

at http://bioinf.cs.ucl.ac.uk/psipred/psiform.html was used and for the discovery of protein repeats the REP program at http://www.embl-heidelberg.de/~andrade/papers/rep/search.html.

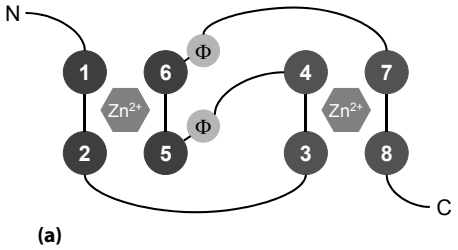
Results and Discussion

FANCL: a really interesting new gene (RING) – Recently the FANCL protein, also known as FAAP43, PHF9 or Pog, has been identified as a key component in the FA core complex carrying the catalytic activity required for the monoubiquitylation of FANCD2 [9]. In that study, FANCL has been described to contain three potential WD40 repeats and a C-terminal PHD-type zinc finger motif, the latter motif being claimed to be responsible for the ubiquitin ligase activity of FANCL. However, PHD finger proteins typically occur in proteins involved in chromatin regulation and no ubiquitin ligase activity has yet been reported for this type of domain [17, 18]. Here, we show that this domain in FANCL contains a slightly atypical RING finger domain.

The PHD (plant homeodomain) type zinc finger and the RING (really interesting new gene) finger are closely related structures that do not resemble classical zinc fingers. Both domains use their Cys/His residues to bind two zinc atoms in a unique "cross-brace" fashion: the first and third pair of metal binding residues bind the first zinc, and the second and fourth pair bind the second zinc (Fig 1A). This zinc ligation is required for proper folding and is essential for the biological activity of both domains. Because of their close resemblance it is difficult to distinguish between PHD and RING fingers and several sequence-based approaches for the classification of complex zinc fingers are in use. The simplest method uses a pattern that relies exclusively on the nature and spacing of the cysteine and histidine residues that ligate to the zinc ions. With this method PHDs are defined by the formula $C\text{-}x_2\text{-}C\text{-}x_{(09\text{-}21)}\text{-}C\text{-}x_{(2\text{-}4)}\text{-}C\text{-}x_{(4\text{-}5)}\text{-}H\text{-}x_2\text{-}C\text{-}x_{(12\text{-}46)}\text{-}C\text{-}x_2\text{-}C$ and RINGs by $C\text{-}x_2\text{-}C\text{-}x_{(09\text{-}39)}\text{-}C\text{-}x_{(1\text{-}3)}\text{-}H\text{-}x_{(2\text{-}3)}\text{-}C\text{-}x_2\text{-}C\text{-}x_{(04\text{-}48)}\text{-}C\text{-}x_2\text{-}C$ where x can be any residue. Clearly, this crude method qualifies FANCL as a PHD finger since metal ligand 4 is a cysteine and metal ligand 5 is a histidine (Fig 1B). A second approach, which relies on the evolutionary relationship within protein families, has recently been suggested to be superior for the discrimination between PHDs and RINGs [17]. This method of protein classification is also used by several databases such as Pfam [19] and has been proven to be highly successful for the prediction of protein functionality. In order to classify FANCL, we performed a PSI-BLAST database search [20] and found that the C-terminal domain of FANCL had weak but statistically significant homologies with several other proteins (Fig 1B). Interestingly, all of these proteins are classified as RINGs by Pfam indicating that the motif in FANCL should be designated as a RING rather than as a PHD finger.

An additional observation favoring the RING comes from the hydrophobic core of the motif. Although both RING and PHD domains have conserved hydrophobic residues at the same positions, PHDs have an additional hydrophobic tryptophan residue two positions N-terminal of the 7th metal ligand (Fig. 1B). This tryptophan, which is highly conserved throughout the PHD family, is an essential part of the hydrophobic core of PHDs [21]. This important residue is missing in FANCL (Fig. 1B), which indicates that FANCL is not a PHD protein. In RING structures reported, the residue that occurs at this position is solvent-exposed and is not part of the hydrophobic core of the motif [21, 22].

Taken together, FANCL contains a RING finger domain with a slightly atypical Cys/His pattern that is similar to that of PHDs. This odd Cys/His pattern may be the reason why Meetei *et al*. [9] have previously classified FANCL as a PHD, but this pattern is of only minor importance since RINGs are flexible with regard to the ligating residue, e.g. His can replace Cys and even Asp can be used as a Zn^{2+} ligand [21]. Overall, the RING domain of FANCL closely resembles that of the RING consensus and has a hydrophobic core that is typical for that of RINGs.



(a)

```
                1  2                3    4       5  6                        7  8
PHF9_HUMAN (FANCL)      CGICYAYQLDG--TIPDQV-CDNSQCGQP---FHQICLYEWLRGLLTSRQSFNIIFGECPYC
CG12812_AAF54486        CNICFAYRLDGG-EVPLVS-CDNAKCVLK---CHAVCLEEWFKTLMDGKTFLEVSFGQCPFC
AAO52554_Arabidopsis    CPICYSMFHNGT--IPKFQ-CKT--CKNK---FHAGCIYKWFQT--SHKSN-------CPLC
EAK91327_Candida        CAICYSILHQ-DSSLPSKT-CTT--CSNK---FHAACLYKWFKS--SGSST-------CPLC
CAA88657_Saccharomyces  CAICYSILHAVDRKLPSKT-CPT--CKNK---FHGACLYKWFRS--SGNNT-------CPLC
AAU20830_Caenorhabditis CTICMMTVHQQTHQLPKIK-CKQ--CKNK---FHSNCLYKWFES--SNQST-------CPLC
BAD28090_Oryza          CAICLAEFVDG----DTVRVMPV--CGHG---FHARCIERWLAG---GRRS------SCPTC

Pfam consensus zf-C3HC4 CrICleelkepsndfplilpCg...HsGslkyfCrsClerWlssSgnttv.......kCplC
Pfam consensus PHD      CsvCgkpdddaggdllq...CCDg Cdrw...fHlaClgppleeppegk.......WlCpeC
```

(b)

Fig 1. (A) Zinc ligation pattern as found in RING domains. **(B)** Comparison of the RING domains of human FANCL and its *Drosophila* ortholog CG12812 with several other RING domain-containing proteins. The numbers of the zinc ligands are shown above the sequence.

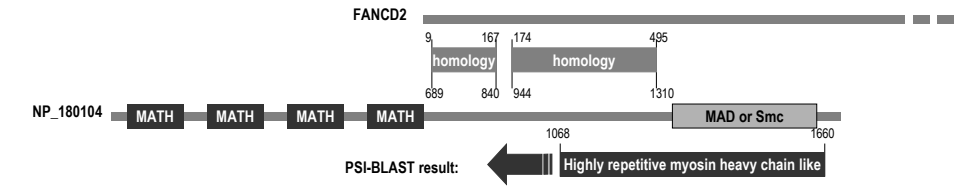
FANCD2: a possible coiled coil protein – In 2001 the gene responsible for the defect in FA-D2 cells was identified encoding a gene product of 1451 residues [5]. Like other FANC proteins it does not contain known protein domains, but similar as the FANCL protein it is more conserved throughout evolution, having orthologs also in non-vertebrate species such as *Drosophila melanogaster*, *Caenorhabditis elegans* and *Arabidopsis thaliana* [5, 23]. In the analysis shown here, evidence is presented indicating that FANCD2 might contain a coiled coil domain similar to that of myosin heavy chain, mitotic checkpoint proteins, or structural maintenance of chromosomes proteins.

The analysis of FANCD2 was started with a PSI-BLAST search since it is more sensitive than BLAST for revealing distant homologs [20]. After the 2nd or further iterations only orthologs of FANCD2 were found, except one 1663 amino acid *A. thaliana* protein which is named 'meprin and TRAF homology domain-containing protein' (Acc. no's: NP_180104, AAD23659 or A84647). Since in this case these ortholog sequences gave little information about the domain structure, the analysis was further focused on the *Arabidopsis* protein. As shown in figure 2A two regions in the N-terminus of FANCD2 are homologous to the central part of the *Arabidopsis* protein, i.e. region 9-167 and 174-495 of FANCD2 with region 689-840 and 944-1310 of the *Arabidopsis* protein, respectively.

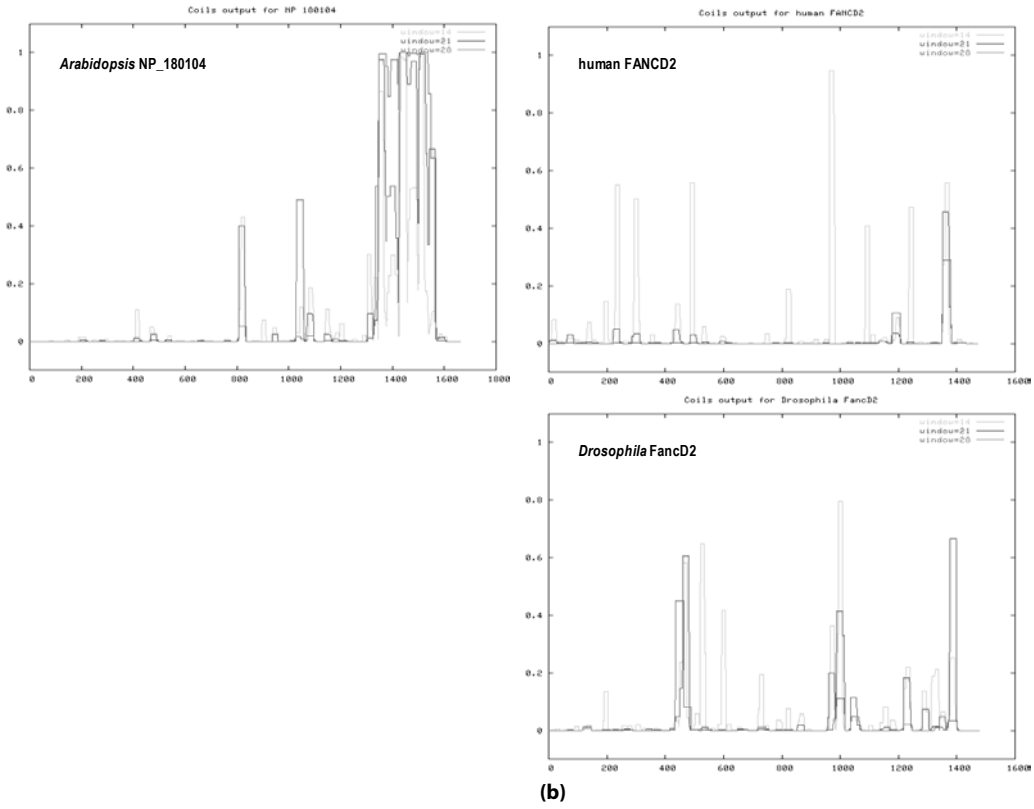
Analysis of the *Arabidopsis* protein using the conserved domain database search (rpsblast at the NCBI website) revealed that four N-terminal MATH domains are present (Fig 2A). The MATH domain (meprin and TRAF homology) is a domain present in both extracellular mammalian tissue-specific endopeptidases (meprins) and in intracellular proteins interacting with TNF receptors (TRAF proteins). Since the homology between FANCD2 and the *Arabidopsis* protein is not based on the presence of a MATH domain in FANCD2 (Fig 2A), this domain is not further discussed here. At the C-terminus, the rpsblast tool predicted either a MAD or Smc domain but both were only partially present, with 26% aligned (E value 9×10^{-6}) and 20% aligned (E value 3×10^{-11}), respectively. The MAD domain (mitotic arrest-deficient) is present in several mitotic checkpoint proteins such as yeast Mad1p that monitor proper attachment of the bipolar spindle to the kinetochores of aligned sister chromatids and causes a cell cycle arrest when failure occurs. In addition, the Smc domain (structural maintenance of chromosomes) is present in chromosome segregation ATPases. An interesting member of this family is the *S. pompe* Rad18 protein that acts in a DNA repair pathway for removal of UV-induced DNA damage and ionizing radiation damage that is distinct from classical nucleotide excision repair pathway and is probably also involved in the maintenance of chromatin structure.

Like the MATH domain, the homology between FANCD2 and the *Arabidopsis* protein does not seem to be based on the presence of a MAD/Smc-like domain in FANCD2: only the central part *between* the four N-terminal MATH domains and the C-terminal MAD/Smc-like domain of the *Arabidopsis* protein is homologous to FANCD2. However, further analysis of the *Arabidopsis* protein using PSI-BLAST revealed that the C-terminal MAD/Smc-like domain is extended towards the N-terminus (Fig 2A). This region is highly repetitive and homologous to myosin heavy chain-containing proteins. Although this region appears to start at residue 1068, upon further iterations using the PSI-BLAST program it appears to be extended even more towards the N-terminus, as indicated by the arrow in figure 2A. This result shows that myosin heavy chain and MAD/Smc-like domains are related and, more importantly, that FANCD2 may contain a similar domain at its N-terminus. Coiled coil domains are protein-protein interaction motifs consisting of two or more α-helices that twist around each other forming a supercoil. Coiled coil proteins can be both heterodimers or homodimers and usually form 'rod'-like structures providing the cell with cables and networks in the cyto- and nucleocytoskeleton and molecular scaffolds that organize membrane systems and molecular motors [24]. To test for the possible presence of a coiled coil region in both the *Arabidopsis* protein and in FANCD2, the COILS program [25] was used (Fig 2B). In the *Arabidopsis* protein, a high probability of forming a coiled coil region was found in region 1300-1600, which is in agreement with the predicted MAD/Smc-like domain. In addition, the more N-terminal region 800-1300 may also form a coiled coil. This is in agreement with the PSI-BLAST result, although this is predicted with a much lower probability by the COILS program. Interestingly, in the FANCD2 protein coiled coils are predicted throughout the entire protein. Although this prediction is of a low probability, a similar result was obtained with the *Drosophila* Fancd2 ortholog which is only 23% identical to the human sequence (Fig 2B).

In summary, the analysis suggests that the N-terminus of FANCD2, but possibly a more extended region, may form a coiled coil. Although this hypothesis can not be verified until X-ray structural analysis is performed, the possible homo- or heterodimerization characteristic for coiled coils may be investigated by biochemical methods. FANCD2 is not the first example of a DNA repair protein with a coiled coil domain, since other DNA repair proteins such as the *S. pompe* Rad18 proteins also contain coiled coils. Possibly, FANCD2 provides some structural basis that may be required for stabilization of repair intermediates during the handling of DNA interstrand cross-links.



(a)



(b)

Fig 2. (A) Homology between FANCD2 and the *A. thaliana* meprin and TRAF homology domain-containing protein NP_180104. **(B)** Prediction of possible coiled coil regions using the COILS program. The COILS program compares a protein sequence to a databases of known coiled-coils and calculates the probability that the protein region will adapt a coiled coil conformation. In both human and *Drosophila* FANCD2 coiled coils are predicted throughout the entire sequence, in contrast to *Arabidopsis* protein NP_180104 in which only C-terminal coiled coils are predicted.

FANCA and FANCC: similar folds as HEAT repeat-like proteins – Since standard BLAST searches or other pairwise sequence comparison algorithms often fail to detect similarities when identities fall below ~30%, enhanced database searching algorithms have been developed, such as PSI-BLAST that make use of profiles of related sequences in stead of single sequences. Unfortunately, applying this more sensitive way of searching also failed for most of the FANC proteins and these proteins have therefore always been classified as 'orphan' proteins. Although it may not be possible to classify the FANC proteins using PSI-BLAST, we hypothesized that methods that incorporate structural information during the searching process may give information about type of fold of the FANC proteins. Unlike the number of protein sequences, the number of different folds in the protein universe is rather small (estimated at 1000-8000). This implies that although the FANC proteins have unique sequences, they are likely to have similar structures as other proteins in the database. To test this, the mGenThreader server was used which is a modified version of the original GenThreader fold recognition method [26]. This program makes use of PSI-BLAST profile and secondary structure prediction as inputs to enhance sensitivity. In brief, this program predicts whether the sequence and the predicted secondary structure of the query has significant homology to one or more entries from the PDB database, which contains all known protein structures.

To test the reliability of the mGenThreader server, we first tested whether the FANCG and FANCL proteins could be correctly predicted by the program. As discussed previously, FANCG is a tetratricopeptide repeat (TPR) protein [14] and FANCL a RING finger protein (this report). In the PDB database, several structures of proteins containing these domains are present and the mGenThreader algorithm should therefore be able to identify a significant similarity. Interestingly, for both proteins several structures from TPR or TPR-like proteins were found for FANCG at confidence level 'certain' and several structures from RING finger proteins were found for FANCL at medium to low confidence (Table 1). This indicates that results from this server may be reliable, although for FANCG one non-TPR structure was found at confidence level 'certain' indicating that results should still be interpreted with caution.

All other known FANC proteins, FANCA, -B, -C, -D2, -E, and -F, were then submitted to the server. Unfortunately, for FANCB and FANCD2 results were inconsistent, as several different structures were linked to the same protein even though they were predicted at a high confidence level. Results for FANCE and FANCF were only of low confidence and were also inconsistent. However, FANCA and FANCC both showed similarity to several HEAT repeat-containing structures at high confidence without inconsistencies, indicating that FANCA and FANCC could

have similar folds as HEAT repeat-containing proteins. The HEAT repeat (huntingtin, elongation factor 3, A subunit of protein phosphatase 2A, and TOR1) is a unit of a pair of anti-parallel α-helices that is linked by an intraunit loop. HEAT repeat motifs vary in length between 37 and 43 residues and occur in tandem arrays forming superhelical scaffolding matrices that mediate protein-protein interactions but also modulate the conformation and biological behavior of the proteins with which they interact [27]. The HEAT repeat family is related to armadillo/β-catenin-like repeats and it has been noted that many HEAT repeat-containing proteins are involved in intracellular transport processes.

In order to evaluate how well the mGenThreader program performs for HEAT repeat proteins specifically, we tested it for the ATM (Ataxia telangiectasia mutated) protein. ATM is a large 370 kDa protein with a C-terminal PI3K-like (Phosphatidylinositol 3-kinase) domain. Outside its highly conserved C-terminal domain ATM is, similarly as the FANC proteins, only sparsely conserved. In addition, this region can not be classified into a known protein domain using PSI-BLAST or Pfam and no repeats are detected by the REP program at a statistically significant level. However, an elaborate study by Perry and Klecker [28] has found that the nonkinase portions of ATM and ATM-related proteins such as ATR and TOR are in fact composed of many HEAT repeats. As expected, the mGenThreader program predicts a similarity between ATM and several structures containing HEAT repeats, indicating that the results for FANCA and FANCC should be reliable.

In conclusion, FANCA and FANCC may have HEAT repeat-like 3D structures. The putative HEAT repeats in FANCA and FANCC may have a similar function as the HEAT repeats of ATM or ATR. The HEAT repeats of ATM fold into two main domains comprising a head and an arm [29]. Upon DNA binding of ATM, the conformation of the arm domain changes and wraps around the double helix and forms a ring around the DNA that may be reminiscent of the 'sliding clamp' structures such as PCNA [30]. It can be hypothesized that the putative HEAT-like repeats in FANCA and FANCC may also be involved in DNA binding and recognize specific structures, such as stalled replication forks.

Table 1. mGenThreader results for FANCA, FANCC, FANCG and FANCL

Query	PDB Id	E value (Confidence)	Name	Protein (SwissProt)	Domain(s)	Remarks
FANCA (1-700)	1qgr	9.10^{-4} (certain)	Structure of importin beta bound to the ibb domain of importin alpha	IMB1_HUMAN IMA2_HUMAN	HEAT repeat ARM repeat	
	1b3u	0.009 (high)	Scaffold protein. Crystal structure of constant regulatory domain of human pp2a, pr65alpha	2AAA_HUMAN	HEAT repeat	
FANCA (701-1455)	1qgr	0.002 (high)	Structure of importin beta bound to the ibb domain of importin alpha	IMB1_HUMAN IMA2_HUMAN	HEAT repeat ARM repeat	
	1b3u	0.014 (medium)	Scaffold protein. Crystal structure of constant regulatory domain of human pp2a, pr65alpha	2AAA_HUMAN	HEAT repeat	
FANCC	1qgr	5.10^{-4} (certain)	Structure of importin beta bound to the ibb domain of importin alpha	IMB1_HUMAN IMA2_HUMAN	HEAT repeat ARM repeat	
	1b3u	0.004 (high)	Scaffold protein. Crystal structure of constant regulatory domain of human pp2a, pr65alpha	2AAA_HUMAN	HEAT repeat	
	3bct	0.008 (high)	The armadillo repeat region from murine beta-catenin	CTNB_MOUSE	ARM repeat	
FANCG	1fch	3.10^{-5} (certain)	Crystal structure of the pts1 complexed to the tpr region of human pex5	PEX5_HUMAN	TPR repeat	
	1qqe	4.10^{-5} (certain)	Crystal structure of the vesicular transport protein sec17	SC17_YEAST	Aromatic-di-Alanine (AdAR) repeat: Similar structure and member of the TPR superfamily	
	1qsa	4.10^{-5} (certain)	Crystal structure of the 70 kda soluble lytic transglycosylase slt70 from escherichia coli at 1.65 angstroms resolution	SLT_ECOLI	Transglycosylase SLT domain	False positive
FANCL	1v87	0.072 (medium)	Solution structure of the ring-h2 finger domain of mouse deltex protein 2	DTX2_MOUSE	RING finger	
	1chc	0.331 (low)	Equine herpes virus-1 (c3hc4, or ring domain) (nmr, 1 structure)	ICP0_HSVEB	RING finger	
	1jm7	0.394 (low)	Solution structure of the brca1/bard1 ring-domain heterodimer	BRC1_HUMAN BAR1_HUMAN	Both RING finger	

Note: Per query only a few hits are shown. The FANCA sequence was too large to be submitted at once to the server.

Conclusion

For 4 of the FANC proteins analyzed, i.e. FANCL, -D2, -A, and -C, relevant information was obtained about their domain structures. Unfortunately, no clues were found for FANCB, FANCE, and FANCF. However, as threading software, used for FANCA and FANCC, will improve and a more diverse set of structures will become available in the PDB database, these proteins might become classifiable as well. FANCL is a member of the RING finger family of proteins, which is in agreement with previous experimental data. The predictions for FANCD2, FANCA, and FANCC, however, are less certain but may provide useful starting points for further biochemical studies. As a putative coiled coil protein, the possible homo- or heterodimerization of FANCD2 could be tested. The putative HEAT-like repeats may be involved in DNA binding and the binding of the FA core complex to specific DNA structures such as stalled replication forks may be experimentally verifiable.

References

1 Lo Ten Foe JR, Rooimans MA, Bosnoyan-Collins L, Alon N, Wijker M, Parker L, Lightfoot J, Carreau M, Callen DF, Savoia A, Cheng NC, van Berkel CG, Strunk MH, Gille JJ, Pals G, Kruyt FA, Pronk JC, Arwert F, Buchwald M, Joenje H (1996). Expression cloning of a cDNA for the major Fanconi anemia gene, FAA. *Nat Genet* **14:** 320–323.

2 The Fanconi anemia/breast cancer consortium (1996). Positional cloning of the Fanconi anaemia group A gene. *Nat Genet* **14:** 324–328.

3 Meetei AR, Levitus M, Xue Y, Medhurst AL, Zwaan M, Ling C, Rooimans MA, Bier P, Hoatlin M, Pals G, de Winter JP, Wang W, Joenje H (2004). X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet* **36:** 1219–1224.

4 Strathdee CA, Gavish H, Shannon WR, Buchwald M (1992). Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **356:** 763–767.

5 Timmers C, Taniguchi T, Hejna J, Reifsteck C, Lucas L, Bruun D, Thayer M, Cox B, Olson S, D'Andrea AD, Moses R, Grompe M (2001). Positional cloning of a novel Fanconi anemia gene, *FANCD2*. *Mol Cell* **7:** 241–248.

6 de Winter JP, Léveillé F, van Berkel CG, Rooimans MA, van Der Weel L, Steltenpool J, Demuth I, Morgan NV, Alon N, Bosnoyan-Collins L, Lightfoot J, Leegwater PA, Waisfisz Q, Komatsu K, Arwert F, Pronk JC, Mathew CG, Digweed M, Buchwald M, Joenje H (2000). Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am J Hum Genet* **67:** 1306–1308.

7 de Winter JP, Rooimans MA, van Der Weel L, van Berkel CG, Alon N, Bosnoyan-Collins L, de Groot J, Zhi Y, Waisfisz Q, Pronk JC, Arwert F, Mathew CG, Scheper RJ, Hoatlin ME, Buchwald M, Joenje H (2000). The Fanconi anaemia gene *FANCF* encodes a novel protein with homology to ROM. *Nat Genet* **24:** 14–16.

8 de Winter JP, Waisfisz Q, Rooimans MA, van Berkel CG, Bosnoyan-Collins L, Alon N, Carreau M, Bender O, Demuth I, Schindler D, Pronk JC, Arwert F, Hoehn H, Digweed M, Buchwald M, Joenje H (1998). The Fanconi anaema group G gene *FANCG* is identical with *XRCC9*. *Nat Genet* **20:** 281–283.

9 Meetei AR, de Winter JP, Medhurst AL, Wallisch M, Waisfisz Q, van de Vrugt HJ, Oostra AB, Yan Z, Ling C, Bishop CE, Hoatlin ME, Joenje H, Wang W (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35:** 165–170.

10 Levitus M, Rooimans MA, Steltenpool J, Cool NF, Oostra AB, Mathew CG, Hoatlin ME, Waisfisz Q, Arwert F, de Winter JP, Joenje H (2004). Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes. *Blood* **103:** 2498–2503.

11 Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J, Grompe M, D'Andrea AD (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7:** 249–262.

12 Thomashevski A, High AA, Drozd M, Shabanowitz J, Hunt DF, Grant PA, Kupfer GM (2004). The Fanconi anemia core complex forms four complexes of different sizes in different subcellular compartments. *J Biol Chem* **279:** 26201–26209.
13 Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C, Persky N, Grompe M, Joenje H, Pals G, Ikeda H, Fox EA, D'Andrea AD (2002). Biallelic inactivation of *BRCA2* in Fanconi anemia. *Science* **297:** 606–609.
14 Blom E, van de Vrugt HJ, de Vries Y, de Winter JP, Arwert F, Joenje H (2004). Multiple TPR motifs characterize the Fanconi anemia FANCG protein. *DNA Repair* **3:** 77–84.
15 Bork P, Blomberg N, Nilges M (1996). Internal repeats in the BRCA2 protein sequence. *Nat Genet* **13:** 22–23.
16 Wong AK, Pero R, Ormonde PA, Tavtigian SV, Bartel PL (1997). RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene BRCA2. *J Biol Chem* **272:** 31941–31944.
17 Scheel H and Hofmann K (2003). No evidence for PHD fingers as ubiquitin ligases. *Trends Cell Biol* **13:** 285-287.
18 Aravind L, Iyer LM and Koonin EV (2003). Scores of RINGS but no PHDs in ubiquitin signaling. *Cell Cycle* **2:** 123–126.
19 Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR (2004). The Pfam protein families database. *Nucleic Acids Res* **32:** D138–D141.
20 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25:** 3389–3402.
21 Capili AD, Schultz DC, RauscherIII FJ, Borden KLB (2001). Solution structure of the PHD domain from the KAP-1 corepressor: structural determinants for PHD, RING and LIM zinc-binding domains. *EMBO J* **20:** 165–177.
22 Borden KL (1998). RING fingers and B-boxes: zinc-binding protein-protein interaction domains. *Biochem Cell Biol* **76:** 351–358.
23 Castillo V, Cabré O, Marcos R and Surrallés J (2003). Molecular cloning of the *Drosophila* Fanconi anaemia gene FANCD2 cDNA. *DNA Repair* **2:** 751–758.
24 Rose A, Meier I (2004). Scaffolds, levers, rods and springs: diverse cellular functions of long coiled-coil proteins. *Cell Mol Life Sci* **61:** 1996–2009.
25 Lupas A, Van Dyke M, Stock J (1991). Predicting coiled coils from protein sequences. *Science* **252:** 1162–1164.
26 Jones DT (1999). GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol* **287:** 797–815.

27 Groves MR, Barford D (1999). Topological characteristics of helical repeat proteins. *Curr Opin Struct Biol* **9:** 383–389.
28 Perry J, Kleckner N (2003). The ATRs, ATMs, and TORs are giant HEAT repeat proteins. *Cell* **112:** 151–155.
29 Llorca O, Rivera-Calzada A, Grantham J, Willison KR (2003). Electron microscopy and 3D reconstructions reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA. *Oncogene* **22:** 3867–3874.
30 Ellison V, Stillman B (2001). Opening of the clamp: an intimate view of an ATP-driven biological machine. *Cell* **106:** 655–660.

CHAPTER

6

Towards a *Drosophila melanogaster* model for the Fanconi anemia pathway

Fanconi anemia (FA) is a recessively inherited genome instability syndrome characterized by at least 11 complementation groups. To date, 9 FA genes have been identified and most of their gene products assemble in a multisubunit core complex which is required to activate the downstream FA protein FANCD2 by monoubiquitylation. This pathway has evolved relatively recently during evolution, as only vertebrates posses a full set of FA genes. However, the fruit fly *Drosophila melanogaster* possesses two FA genes: the E3 ubiquitin ligase dmFANCL and its target protein dmFANCD2 indicating that this non-vertebrate may have a 'basic' version of the FA pathway, whose characterization may provide novel clues for understanding the pathway in vertebrates. In a pilot mass spectrometry experiment, dmFANCL was found to interact with the CAF-1 p105 protein, possibly linking the FA pathway to chromatin assembly processes. In addition, preliminary results indicate that *Drosophila* cells only express the activated dmFANCD2-L isoform, suggesting that the fruit fly FA pathway could be characterized as 'always on', in contrast to mammalian cells where the activation of FANCD2 seems to be highly regulated.

Eric Blom, Annette L. Medhurst, Roel C. van der Schors, Johan P. de Winter, Fré Arwert & Hans Joenje

Preliminary results

Introduction

Fanconi anemia (FA), a chromosomal instability syndrome characterized by developmental abnormalities, progressive pancytopenia and cancer predisposition, is genetically a highly heterogeneous disorder. Cell fusion experiments have revealed the existence of at least 11 complementation groups (FA-A, B, C, D1, D2, E, F, G, L, I, and J) and the corresponding genes of 9 complementation groups have been identified (*FANCA, B, C, D1, D2, E, F, G,* and *L*) [1-10]. Biochemical studies have shown that FANC proteins function in a single pathway and assemble in a large protein complex [11] that is essential for activation of the downstream FANCD2 protein by monoubiquitylation at Lys^{561} [5]. This activation is a critical step required to allow further downstream reactions, such as the association of FANCD2 with BRCA1 in nuclear repair foci [12]. In spite of a considerable progress in the FA field [13, 14], the exact function of monoubiquitylated FANCD2 and how this protein protects the genome against DNA damage such as cross-links is currently not well understood.

In this study, we hypothesize that the fruit fly *Drosophila melanogaster* may be a helpful model for studying the FA pathway. In contrast to vertebrates, the non-vertebrate *Drosophila melanogaster* possesses only two *FANC* genes, suggesting that *Drosophila* might have a 'basic' version of the FA pathway. It consists of the E3 ubiquitin ligase dmFANCL and its substrate, dmFANCD2. Since no other core complex proteins are present in *Drosophila*, it is unclear how the dmFANCL protein can perform its task without the help of the core complex. In the mammalian system, FANCL can not monoubiquitylate FANCD2 in the absence of a functional core complex. It was hypothesized that other proteins may function as 'functional analogs' in *Drosophila* and replace the function of the vertebrate core complex. Therefore, the identification of these putative analogs may be valuable for understanding the nature and function of the human core complex.

Our results show that both dmFANCL and dmFANCD2 are expressed at the mRNA level in the commonly used *Drosophila* S2 cell line [15]. Currently, antibodies for these proteins are not available and the expression at the protein level could therefore not yet be studied. Since large numbers of S2 cells can easily be obtained, this cell line is an excellent tool for mass spectrometry experiments for which large amounts of cells are required. In a pilot mass spectrometry experiment, dmFANCL was found to interact with the chromatin assembly factor 1 subunit p105 protein (CAF-1 p105). Moreover, preliminary results using a myc-tagged dmFANCD2 construct suggest that dmFANCD2 is only present in its monoubiquitylated form (dmFANCD2-L) and no dmFANCD2-S is observed. This differs from the situation in mammalian cells in which both forms are expressed with the FANCD2-S form being

predominant. If this observation can be verified by further experiments, this may indicate that the FA pathway in *Drosophila*, in contrast to vertebrates, is to be characterized as 'always on'.

Materials and methods

Cell culture and transfection procedure – Drosophila melanogaster S2 cells [15], a kind gift from Dr. David Baker (Leiden University Medical Center, The Netherlands), were grown in Schneider's insect medium (Sigma) supplemented with 10% heat-inactivated bovine serum (Gibco) and 5 ml/L Penicillin-Streptomycin (Invitrogen) without CO_2 at 23°C, in closed flasks. To stably transfect cells, exponentially growing cells were seeded in T-25 cm^2 flasks at a concentration of 1×10^6 cells/ml and after 24 hrs were transfected using a calcium phosphate transfection kit (Invitrogen) according to manufacturer's instructions. For stable expression of a single construct, 20 μg plasmid DNA of the expression vector (pAc5.1 + insert) and 1 μg DNA of the selection vector (pCoHygro) was used. For stable expression of two different constructs, we used 10 μg plasmid DNA of each expression vector and only 0.75 μg of the pCoHygro plasmid. After 72 hrs, the cells were centrifuged, resuspended in fresh medium and replated into the same flask and after 96 hrs cells were selected in medium containing 300-400 μg/ml Hygromycin B (Roche or Invitrogen) until resistant cells appeared (typically after 2 weeks).

Constructs – A cDNA clone containing full-length *Drosophila melanogaster dmFANCD2* was a kind gift from Dr. J. Surrallés (Universitat Autònoma de Barcelona, Spain). Since no antibodies were available for the dmFANCD2 protein, a N-terminal c-myc tag was added and cloned into the pAc5.1 expression vector (Invitrogen). Since the cDNA was not available for *dmFANCL* and technical difficulties impeded the cloning of the cDNA, the genomic sequence was obtained by PCR on genomic DNA isolated from S2 cells. An N-terminal flag-tag was then added to the first exon and the construct was cloned into the pAc5.1 expression vector. It was expected that S2 cells were able to correctly splice this construct because the intron sequences were left intact. In the human system, for both FANCD2 and FANCL proteins an N-terminal tag is known not to interfere with protein functionality. In addition, a C-terminal V5- and His-tagged construct that contained exon 2 of *dmFANCL* (encoding the majority of the protein, see figure 1) was prepared in the same vector. This construct was used in the mass spectrometry experiment.

RT-PCR dmFANCL and dmFANCD2 – Total RNA was extracted from S2 cells using TRIzol reagent (Invitrogen). First-strand cDNA synthesis was then performed using SuperScript II reverse transcriptase (Invitrogen) according the manufacturer's instructions. An aliquot of the synthesized cDNA was then amplified using a standard PCR reaction with an annealing temperature of 58°C and 35 cycles on a Robocycler 96 (Stratagene). The following primers were used: *dmFANCL-F:* TCAGCCAAGTGGAGCGTGCAT, *dmFANCL-R:* CACTTGCAGGCAGTACTCGTT, *dmFANCD2-F:* GAGCGAGTGGAGTGATAAGT, and *dmFANCD2-R:* GCTTTGTCCTTCGTATTCAG.

Mass spectrometry experiment – S2 cells were stably transfected with the C-terminal V5- and His-tagged dmFANCL-exon2 plasmid. As a negative control, S2 cells were transfected with a LacZ plasmid with identical C-terminal V5- and His-tags. Approximately 5×10^9 cells were propagated and a nuclear extract was prepared using a similar procedure as previously described [16, 17]. The nuclear pellet was then lysed in 40 ml NP-40 lysisbuffer containing 10% glycerol (50 mM Tris-HCl, pH7.5, 150 mM NaCl, 1% NP-40, 10% glycerol, 500 μg/ml Pefabloc, 1 μg/ml apoprotinin, 1 μg/ml pepstatin, 1 μg/ml leupeptin) and sheared through a 19-gauge needle. The lysate was then clarified by centrifugation at 4°C, after which α-V5 and α-His(C-term.) antibodies (both from Invitrogen) were added and incubated overnight at 4°C with continuous mixing. The following day, the immune complexes were collected by adding protein A sepharose CL-4B (Amersham Biosciences) for 2 hours at 4°C with mixing. The beads were then washed with lysis buffer and the samples were run on a standard 10% polyacrylamide SDS gel electrophoresis (Hoefer SE 400) gel system. The gel was fixed in 50% ethanol/3% phosphoric acid overnight, washed 3 times for 15 min in MilliQ water, and stained with Coomassie brilliant blue G-250 solution (Sigma) for 24 hours. Bands were excised from the gel and analyzed using mass spectrometry. Note that in this type of experiment the amount of lysis buffer was critical for the observed level of background. When too small a volume is used, the protein concentration becomes too high causing an aspecific polymerization and precipitation of proteins. These precipitated proteins become trapped within the beads and can not be removed during the washing step of the beads [18]. The experiment was first optimized with fewer cells and the gels were stained with the more sensitive silver-staining method instead (Invitrogen SilverQuest staining kit).

Immunoblotting and coimmunoprecipitation – Immunoblotting samples were lysed in 2× sample buffer (80 mM Tris-HCl pH 6.8, 0.8% sodium dodecyl sulfate, 100 mM dithiothreitol (DTT), 4% glycerol, 0.4% bromphenolblue), boiled for 5 min and subjected to standard polyacrylamide SDS gel electrophoresis. After transfer to an Immobilon-P membrane (Millipore), α-V5 (1:5000, Invitrogen) was used as primary antibody and goat α-mouse-HRP (1:2000, DAKO) as secondary antibody after which the blot was developed using the ECL Western blotting system (Amersham Biosciences).

For coimmunoprecipitation total cell extracts, prepared in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% NP-40, 500 μg/ml Pefabloc, 1 μg/ml apoprotinin, 1 μg/ml pepstatin, 1 μg/ml leupeptin), were incubated with α-V5 antibody or with α-flag M2 resin (Sigma) 2 h or overnight, at 4°C. When α-V5 antibody was used, protein A beads were added afterwards and incubated at 4°C for 30 min. After washing the beads in lysis buffer, 2× sample buffer was added and boiled for 5 min. The samples were then run on a standard polyacrylamide SDS gel electrophoresis and analyzed by immunoblotting.

Monoubiquitylation of dmFANCD2 – S2 cells, stably transfected with either flag-tagged dmFANCL and myc-tagged dmFANCD2 or myc-tagged dmFANCD2 alone, were seeded at a concentration of approximately of 1×10^6 cells/ml in the presence of 1 or 2 μM final concentration mitomycin C (MMC) or 2 mM final concentration hydroxyurea (HU) or cells were gassed with O_2 or N_2. The following day, 0.5 ml cell suspension was collected and washed once in PBS after which the cell pellet was lysed in 2× sample buffer, then heated at 95°C for 5 min, followed by a standard 6.5% SDS gel electrophoresis after which the proteins were transferred to an Immobilon-P transfer membrane (Millipore) in transfer buffer (25 mM Tris-HCl, 200 mM glycine, 15% methanol) at ~80 mA overnight at 4°C. After blocking in 5% (w/v) non-fat dry milk in TBST (10 mM Tris-HCl, 150 mM NaCl, 0.2% Tween-20) the blot was incubated in primary antibody (α-myc, Invitrogen), washed 3 times in TBST, then incubated in secondary antibody (goat α-mouse HRP, DAKO) and washed 5 times in TBST after which the ECL Western blotting analysis system (Amersham Pharmacia) was used to develop the blot.

Results

Evolutionary conservation of dmFANCL – Unlike other FA core complex members, FANCL is conserved in the non-vertebrate *D. melanogaster*. *Drosophila* FANCL is a 381 residue protein named CG12812 in the database and no experimental studies have yet been published. Overall, the protein is poorly conserved with a percentage identity of only 22% and a percentage similarity of 35%, which is close to the detection limit of database search programs. The RING domain, however, is much more conserved (Fig. 1) suggesting that it has maintained its ubiquitin ligase activity during evolution. In order to exclude the possibility that the protein CG12812 is just a random protein sharing a similar C-terminal RING domain, a PSI-BLAST search was performed using the *Drosophila* sequence *without* the RING domain. As expected, statistical significant homologies between dmFANCL, human FANCL and several other vertebrate FancL sequences were found. In addition, no statistical significant homology was found to any other protein in the database, verifying that the sequences found are true orthologs. Furthermore, although suggested by Meetei *et al.* [9] no evidence was found for the presence of possible WD40 repeats in the N-terminus in our analysis. On the genomic level, *dmFANCL* is encoded by 3 exons with 2 small introns of about 55 bp each. The central exon 2 encodes the majority of the protein. The exon boundaries are indicated in figure 1.

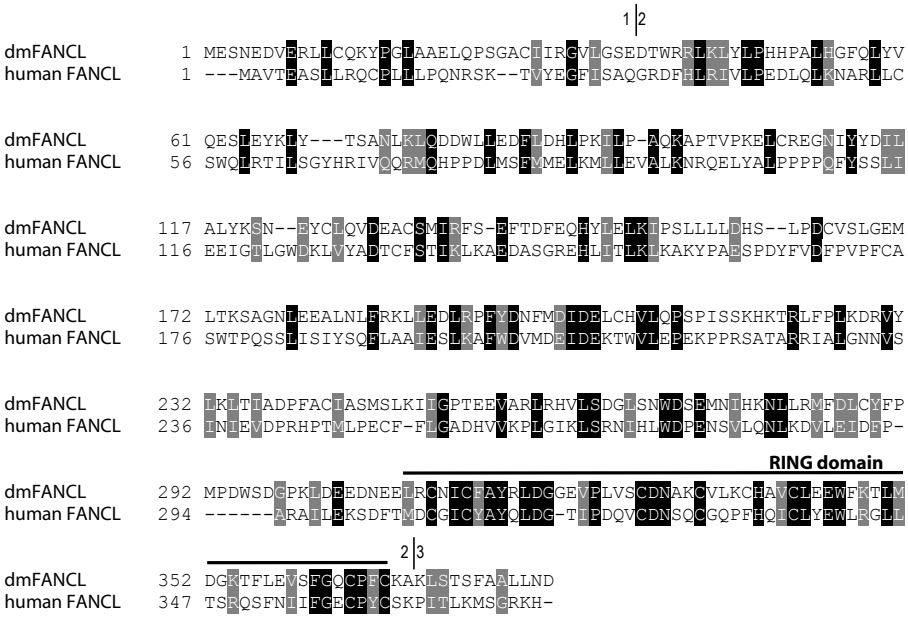


Fig. 1. Comparative sequence analysis of dmFANCL and its human ortholog. Identical residues are indicated in black and similar residues in gray (BoxShade program). Exon boundaries and the RING domain are indicated on top of the sequence.

Expression of dmFANCL and dmFANCD2

In order to determine whether *dmFANCL* and *dmFANCD2* are expressed in *Drosophila* S2 cells, RT-PCR was performed on total RNA with primers that anneal in different exons. Despite a considerable genomic DNA contamination in the RNA preparation, it was evident that for both *dmFANCL* and *dmFANCD2* smaller PCR fragments derived from mRNA were present (Fig. 2). This indicates that *dmFANCL* and *dmFANCD2* are expressed in S2 cells and that this cell line may be used as a tool for studying the FA pathway in *Drosophila*. Since no antibodies were available, expression of these gene products on the protein level could not be studied.

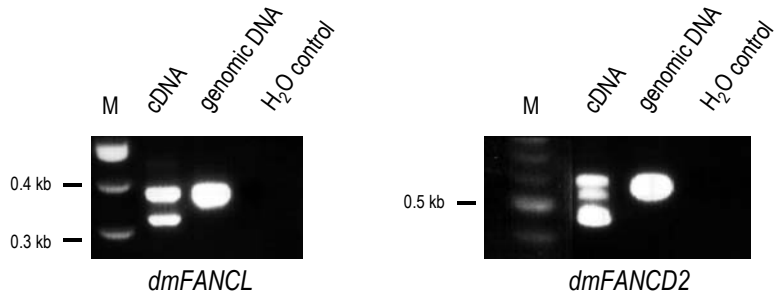


Fig. 2. Expression of dmFANCL and dmFANCD2 mRNAs in the Drosophila melanogaster S2 cell line as shown by RT-PCR technique.

Pilot experiment mass spectrometry of dmFANCL-interacting proteins – S2 cells were stably transfected with C-terminal V5- and His-tagged *dmFANCL-exon2* plasmid that expressed almost the entire protein (Fig. 1). This construct was chosen because a full-length cDNA clone was not available. In addition, the α-V5 antibody performed very well on direct western blot and in immunoprecipitation experiments showing a strong signal without any background, indicating this antibody is a suitable tool for the intended experiment. As a negative control, the cell line was stably transfected with a plasmid containing LacZ with same C-terminal V5- and His-tags. After immunoprecipitation, running and staining the gel, several differences between the negative control and *dmFANCL-exon2* construct were evident indicating that several proteins were interacting specifically to the two dmFANCL constructs. The pattern observed was identical between independent experiments; a representative experiment is shown in figure 3. Protein bands that were identical in both lanes were considered as proteins that aspecifically bind to the beads. Four bands that were different for the *dmFANCL-exon2* construct in the range of 70-110 kDa were excised and analyzed by mass spectrometry (Fig 3). Band 3 and 4 were both identified as heat shock proteins suggesting that these are false positives

possibly caused by the fact that the constructs were overexpressed. In addition, band 2 could not be analyzed due to a low signal strength. Band 1 was identified as chromatin assembly factor CAF1-p105. Although this identification was at a low confidence level due to a low signal strength, the size of the excised band (~110 kDa) matches the predicted size of the protein identified (105 kDa) indicating that this result could be reliable.

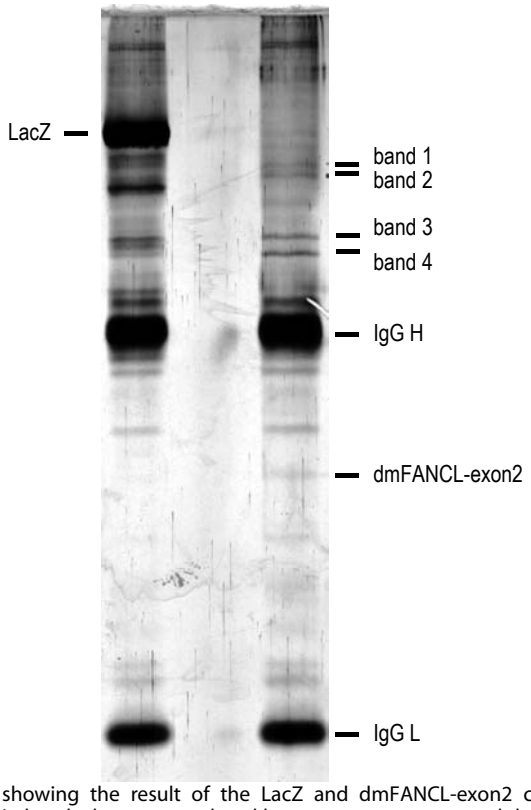


Fig. 3. Silver-stained gel showing the result of the LacZ and dmFANCL-exon2 coimmunoprecipitation experiment. The four protein bands that were analyzed by mass spectrometry and the immunoglobin heavy and light chains (IgG H and IgG L) are indicated.

Protein interaction network of dmFANCL – In an elaborate study, Giot *et al.* [19] recently published a two-hybrid-based protein interaction map of the fruit fly proteome. From a total of 10,623 predicted transcripts, 7048 proteins were found to interact in 20,405 different interactions and this draft map was later refined to a higher confidence map of 4679 proteins and 4780 interactions using computational methods. In their dataset, 8 interactions are present for dmFANCL (table 1) and none for dmFANCD2. A protein interaction map based on the refined dataset is shown in figure 4. From the 8 interactions 2 interacting proteins, a novel E2 ligase and the DNA polymerase α-73 [20], seemed interesting and were therefore experimentally tested using the different coimmunoprecipitation technique. Both cDNAs were cloned by PCR and a C-terminal V5- and His-tagged plasmid was prepared for both (pAc5.1 vector). This plasmid was stably cotransfected with a N-terminal flag-tagged dmFANCL(genomic) construct in S2 cells. Unfortunately, no interaction was found using an α-V5 immunoprecipitation experiment although both constructs were expressed at high levels (Fig. 5). In addition, in a mammalian-two-hybrid pilot experiment with human FANCL and the human counterparts for the E2 ligase and DNA polymerase also no interactions were found (data not shown). The other 6 interacting proteins from the dataset have not been tested to date.

Table 1. *Drosophila* dmFANCL-interacting proteins found by Giot *et al.* [19]

		Confidence	Remarks
CG10293 (how)	CG12812 (dmFANCL)	low	held out wings 405 aa, nuclear location, KH domain, The responsible gene for hypomyelinating quaking deficiency, qkI RNA binding, different isoforms
CG11182 (PHDP)	CG12812 (dmFANCL)	low	putative homeodomain protein 220 aa
CG12812 (dmFANCL)	CG1391 (sol)	high	small optic lobes 1597 aa, Calpain family cysteine protease, Zn-finger in Ran binding protein and others
CG12812 (dmFANCL)	CG3166 (aop)	low	anterior open 732 aa, Sterile alpha motif (SAM)/Pointed domain, Ets-domain Negative regulator of photoreceptor development that acts antagonistically to the proneural signal mediated by RAS Tissue specificity: eye
CG12812 (dmFANCL)	CG32206	high	1260 aa, Low-density lipoprotein receptor domain class A, CUB domain, also repeat like structure (?) extracellular protein?
CG12812 (dmFANCL)	CG40127	low	95 aa, ribonuclease, toxic when expressed in E. coli
CG12812 (dmFANCL)	CG5923 (DNApol-α73)	low	653 aa, initial stage of DNA replication in S cerevisiea, phosphorylated in cell cycle dependent manner
CG12812 (dmFANCL)	CG7220	high	154 aa, E2 ligase

Activation of the FA pathway in S2 cells – In the mammalian system, monoubiquitylation of FANCD2 is a hallmark feature of an activated FA pathway. In order to serve as a model for FA, S2 cells should have a similar activation of dmFANCD2. Although this is likely since its E3 ligase dmFANCL is conserved and the monoubiquitylation site is preserved as well [22], this was not yet verified experimentally. An N-terminal myc-tagged dmFANCD2 construct was expressed in S2 cells. In the mammalian system, a N-terminal tag is known not to interfere with protein functionality of FANCD2. Interestingly, under conditions known to induce monoubiquitylation in mammalian cells, such as treatment with mitomycin C or hydroxyurea, only one band was observed (Fig. 6A). Is was hypothesized that the endogenously expressed dmFANCL in S2 cells may not be sufficient to monoubiquitylate the overexpressed myc-dmFANCD2 construct. The cells were therefore cotransfected with the flag-dmFANCL(genomic) construct. Unexpectedly, this cotransfection induced an additional *smaller* band than observed when myc-dmFANCD2 is expressed alone (Fig. 6B). This may indicate that the flag-dmFANCL(genomic) construct may act as an inhibitor instead. When this construct is not spliced, the two introns of about 55 bp each remain in-frame and a slightly larger protein is then encoded, which may act as a dominant negative. The above result may suggest that unlike mammalian cells where FANCD2-S and FANCD2-L are both present, S2 cells express only dmFAND2-L.

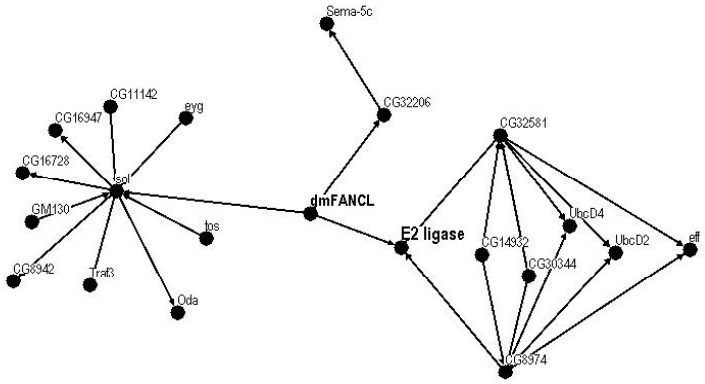


Fig. 4. Protein network of dmFANCL as visualized using Osprey software [21]. All nodes were removed that were more than 2 nodes away from dmFANCL and the putative E2 ligase. Only high confidence interactions are shown. Note that the putative E2 ligase interacts with 3 different E3 ligases: CG32581, CG8974 and dmFANCL. In addition, the E3 ligases CG32581 and CG8974 also interact with three other E2 ligases, UbcD2, UbcD4 and eff, suggesting a redundancy within the network.

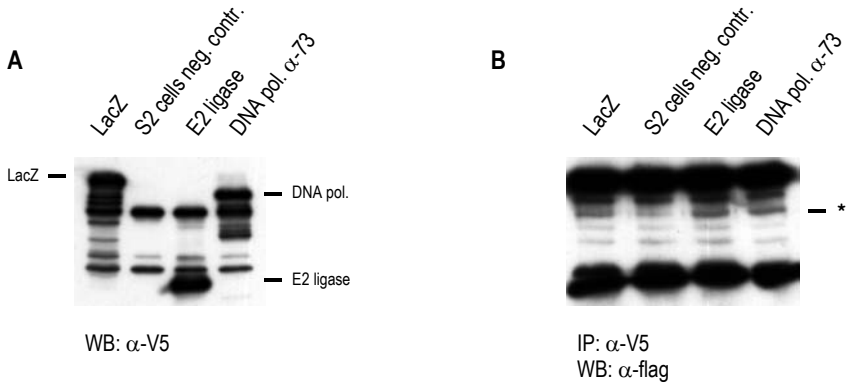


Fig. 5. Coimmunoprecipitation of flag-dmFANCL(genomic) and C-terminal V5-tagged E2 ligase or DNA polymerase. As negative control C-terminal V5-tagged LacZ was used. **(A)** Direct western blot showing that the LacZ, putative E2 ligase, and DNA polymerase α-73 constructs are expressed. **(B)** Immunoprecipitation using α-V5 antibody followed by α-flag M2 Western blot. The position at which the protein expressed from the flag-dmFANCL(genomic) construct was expected is indicated by an asterisk.

Discussion

Although preliminary, the data in this study may give some interesting insight in the FA pathway in *Drosophila* cells. The pilot mass spectrometry experiment suggests that dmFANCL may interact with chromatin assembly factor CAF-1 p105. This is a subunit of a protein complex, named CAF-1, which also contains the p180, p75 and p55 subunits. In human and yeast cells the CAF-1 complex is composed of only three subunits. CAF-1 p105 is a protein of 747 residues with an apparent molecular weight of 105 kDa with several N-terminal WD repeats and a central PEST domain [23]. CAF-1 participates in the assembly of newly synthesized DNA into chromatin during replication but also during DNA repair. Interestingly, CAF-1 interacts with the PCNA component of the DNA replication and repair machinery and both are recruited at sites of DNA lesions and may be involved in sensing of DNA damage [24].

Further optimization of the experiment may reveal several additional proteins that interact with dmFANCL. Increasing the amount of cells may enhance the detection of weaker bands, specifically the ones from smaller proteins. These proteins are less intensively stained by either silver- or Coomassie-staining, biasing the experiment towards the larger proteins. In this experiment, 5×10^9 cells were used, but this number can be increased. This amount corresponds to a culture volume of approximately 400 ml.

Three different proteins are required for ubiquitylation reactions: E1, E2, and E3 ligases. For the FA pathway, the E1 and E2 ligases are currently unknown. It was hypothesized that the E2 ligase found by Giot *et al.* [19] which interacts with dmFANCL using a two-hybrid assay, could be the missing component.

Unfortunately, this interaction was not substantiated by a coimmunoprecipitation experiment, indicating that the interaction may be too weak or too transient to be detected with this technique. A RNAi experiment may therefore give a more decisive clue to this question. Although the two-hybrid interaction between DNA polymerase α-73 and dmFANCL was considered as a low confidence interaction (table 1), this interaction was attempted to confirm using coimmunoprecipitation. In yeast this polymerase is involved in the initial stage of DNA replication [25]. Since activated FANCD2 is expressed during S-phase in mammalian cells, it was speculated that these proteins may be functionally related. Unfortunately, no interaction could be demonstrated.

The monoubiquitylation experiment yielded a highly unexpected result. It was postulated that in *Drosophila* both dmFANCD2-S and dmFANCD2-L are present, similar to what is observed in mammalian cells. However, the results indicated that dmFANCD2-L is expressed suggesting that in *Drosophila* the pathway may be characterized as 'always on'. This suggests that the vertebrate core complex may have been evolved to regulate the monoubiquitylation of FANCD2 in vertebrate cells. Clearly, further experimentation is needed to address this issue more rigorously.

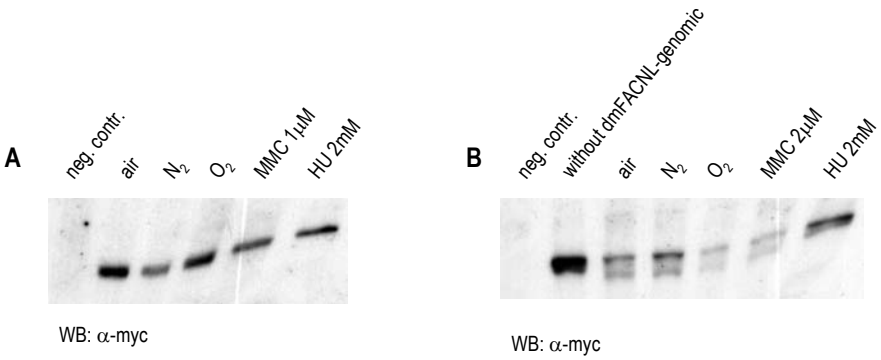


Fig. 6. Monoubiquitylation of N-terminal myc-tagged dmFANCD2 stably expressed in S2 cells. **(A)** Myc-dmFANCD2 construct expressed in S2 cells gives only one band for dmFANCD2. In addition, at the various conditions tested no induction of other isoforms was detected. **(B)** Induction of a smaller additional dmFANCD2 isoform by stably coexpressing the dmFANCL(genomic) construct, which possibly functions as a dominant negative construct (see text). For comparison, the air sample from the experiment of figure A was loaded in the second lane.

References

1 Lo Ten Foe JR, *et al.* (1996). Expression cloning of a cDNA for the major Fanconi anemia gene, *FAA*. *Nat Genet* **14:** 320–323.

2 The Fanconi anemia/breast cancer consortium (1996). Positional cloning of the Fanconi anaemia group A gene. *Nat Genet* **14:** 324–328.

3 Meetei AR, *et al.* (2004). X-linked inheritance of Fanconi anemia complementation group B. *Nat Genet* **36:** 1219–1224.

4 Strathdee CA, Gavish H, Shannon WR, and Buchwald M (1992). Cloning of cDNAs for Fanconi's anaemia by functional complementation. *Nature* **356:** 763–767.

5 Timmers C, *et al.* (2001). Positional cloning of a novel Fanconi anemia gene, *FANCD2*. *Mol Cell* **7:** 241–248.

6 de Winter JP, *et al.* (2000). Isolation of a cDNA representing the Fanconi anemia complementation group E gene. *Am J Hum Genet* **67:** 1306–1308.

7 de Winter JP, *et al.* (2000). The Fanconi anaemia gene *FANCF* encodes a novel protein with homology to ROM. *Nat Genet* **24:** 14–16.

8 de Winter, *et al.* (1998). The Fanconi anaema group G gene *FANCG* is identical with *XRCC9*. *Nat Genet* **20:** 281–283.

9 Meetei AR, *et al.* (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet* **35:** 165–170.

10 Levitus M, *et al.* (2004). Heterogeneity in Fanconi anemia: evidence for 2 new genetic subtypes. *Blood* **103:** 2498–2503.

11 Thomashevski A, *et al.* (2004). The Fanconi anemia core complex forms four complexes of different sizes in different subcellular compartments. *J Biol Chem* **279:** 26201-26209.

12 Garcia-Higuera I, *et al.* (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. *Mol Cell* **7:** 249–262.

13 Joenje H, Patel KJ (2001). The emerging genetic and molecular basis of Fanconi anaemia. *Nat Rev Genet* **2:** 446–457.

14 D'Andrea AD, Grompe M (2003). The Fanconi anaemia/BRCA pathway. *Nat Rev Cancer* **3:** 23–34.

15 Schneider I (1972). Cell lines derived from late embryonic stages of *Drosophila melanogaster*. *J Embyol Exp Morph* **27:** 363–365.

16 Forler D *et al.* (2002). An efficient protein complex purification method for functional proteomics in higher eukaryotes. *Nat Biotechnol* **21:** 89–92.

17 Dignam JD, Lebovitz RM, Roeder RG (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11:** 1475–1489.

18 Harlow E, Lane D (1988). Antibodies, a laboratory manual. Cold Spring Harbor Laboratory. Chapter 11.
19 Giot L *et al.* (2003). A protein interaction map of Drosophila melanogaster. *Science* **302:** 1727–1736.
20 Cotteril S, Lehman IR, McLachlan P (1992). Cloning of the gene for the 73 kD subunit of the DNA polymerase α primase of *Drosophila melanogaster*. *Nucleic Acids Res* **20:** 4325–4330.
21 Breitkreutz BJ, Stark C, Tylers M (2003). Osprey: A network visualization system. *Genome Biol* **4:** R22.
22 Castillo V, Cabre O, Marcos R, Surrallés J (2003). Molecular cloning of the Drosophila Fanconi anaemia gene FANCD2 cDNA. *DNA Repair* **2:** 751–758.
23 Tyler JK *et al.* (2001). Interaction between the *Drosophila* CAF-1 and ASF1 chromatin assembly factors. *Mol Cell Biol* **21:** 6574–6584.
24 Moggs JG *et al.* (2000). A CAF-1–PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. *Mol Cell Biol* **20:** 1206–1218.
25 Uchiyama M, Wang TS (2004). The B-subunit of DNA polymerase α-primase associates with the origin recognition complex for initiation of DNA replication. *Moll Cell Biol* **24:** 7419–7434.

CHAPTER

7

Summary and discussion

Fanconi anemia (FA) is a recessively inherited disorder that was first described in 1927 by the Swiss pediatrician Guido Fanconi. The disease is characterized by aplastic anemia, cancer predisposition (mainly acute myeloid leukemia and squamous cell carcinoma) and diverse congenital anomalies including absent thumbs and radii. Mainly due to the bone marrow failure as well as the cancer risk the mean life expectancy is low, generally 20 years or less. In addition to supportive therapy, bone marrow transplantation is the most important treatment at present.

FA is genetically a highly heterogeneous disease with at least 11 different genes expected to be involved. From these putative 11 genes, 9 have been identified: *FANCA, B, C, D1, D2, E, F, G,* and *L*. Biochemical studies have revealed that FANC proteins function in a single pathway. In this pathway, most FANC proteins (FANCA, B, C, E, F, G, and L) assemble in a nuclear complex – termed the FA core complex – that is required for activation of the FANCD2 protein by monoubiquitylation at Lys^{561}. This activation is essential for further downstream reactions such as the association of FANCD2 with the BRCA1 protein in nuclear repair foci. The FANCD1 gene product, which is identical to the BRCA2 protein, is thought to function downstream of FANCD2 activation and is not required for monoubiquitylation of this protein.

Since FA cells are hypersensitive to agents that induce DNA interstrand cross-links (ICLs), the FA biochemical pathway is thought to be involved in the repair of this type of damage. Interestingly, the DNA interstrand cross-link is one of the most toxic types of DNA damage and if left improperly repaired could result in cell death or cancer. However, the exact role of the FANC proteins in the repair process of ICLs is not well understood. An important reason for this is the fact that the FANC proteins are 'orphan' proteins. In contrast to most other gene products, the FANC proteins are novel proteins that are very limitedly conserved during evolution and do not show any homology to other proteins in the database. This makes an *in silico* prediction of their biochemical function impossible. To investigate this matter, the aim of this thesis was to study the FANC protein sequences in other vertebrates, such as chicken, clawed frog and zebrafish. Amino acids that have remained identical during evolution are likely to be functionally important and in potential could provide important clues about the molecular action of these proteins.

In **chapter 2** the principle of orphan proteins and protein domains is further explained and discussed how a bioinformatic approach could provide important clues about the molecular function of gene products. In a preliminary analysis, FANCG was found to represent a member of the tetratricopeptide repeat (TPR) containing proteins. This was further studied in **chapter 3** and a total of 7 TPRs motifs in FANCG were identified. The TPR motif is an important protein-protein

interaction motif and often plays a role in the assembly of multisubunit protein complexes. Mutation experiments revealed that the TPRs in FANCG are essential for a functional pathway and were confirmed to function as protein-protein interaction modules. In **chapter 4**, the FANCF protein was studied using the approach described in chapter 2 but no protein domain was identified. However, an extensive biochemical study indicated that this protein is likely to function as an adaptor molecule in the FA core complex. The N-terminus is required for binding of the FANCC/FANCE subcomplex and the C-terminus interacts with FANCG. The other FANC protein sequences were studied in **chapter 5**. Evidence was found that FANCA and FANCC are both members of the HEAT-repeat containing proteins and are therefore predicted to be involved as scaffolds in the assembly of the FA core complex, similarly to FANCG. Moreover, the FANCD2 protein is likely to represent a myosin heavy chain-like molecule and FANCL was found to contain a RING domain rather than a PHD domain, as previously reported by others. The presence of a RING domain in FANCL strongly suggests that this protein harbors the catalytic activity required for the monoubiquitylation of FANCD2.

The FA pathway in the fruit fly *Drosophila melanogaster* was studied. Interestingly, the fruit fly genome contains only two *FANC* genes: the upstream catalytic subunit *dmFANCL* and its substrate *dmFANCD2*, suggesting that this species harbors a more elementary FA pathway. It could therefore provide a more simple model for studying FA and may also help to explain the purpose of the multisubunit core complex as found in vertebrates. Preliminary results are presented in **chapter 6**.

The results in this thesis suggest that for many orphan proteins it may be possible to 'de-orphanize' and classify them into a known protein domain family. FANCG was identified as a member of the TPR-containing proteins, FANCA and FANCC are likely HEAT-repeat proteins, and FANCD2 may be a myosin heavy chain-like protein. This suggests that most orphans have arisen by duplication from an ancestral gene, as assumed in chapter 2. An alternative mechanism, the recruitment of randomly occurring open reading frames during evolution, may be therefore be of less importance. Development and application of more sensitive database searching algorithms may result in a decrease of the number of orphans in the database. At present, it is estimated that 10-30% of all predicted open reading frames in fully sequenced genomes are orphans.

Except for FANCL, none of the FANC core complex subunits appears to contain a catalytic domain. Therefore, the overall quaternary structure, possibly in combination with conformational changes during activation of the FA pathway, is likely to represent the most important feature of the FA core complex. There may

exist an important similarity between the FA core complex and the ATM protein. The ATM protein is a large protein of 3056 amino acids that is almost entirely composed of HEAT repeats except a small C-terminal PI3-kinase domain. Single-particle electron microscopy experiments by Llorca *et al.* (2003) revealed that the tertiary structure of ATM is composed of a 'head' and an 'arm' domain. Upon DNA binding of ATM, the conformation of the arm domain changes and wraps around the double helix and forms a ring fixing ATM around the double stranded DNA helix. The FA core complex may function in a similar way: the large quaternary structure of the complex may recognize a particular DNA structure, possibly a stalled replication fork or another DNA repair intermediate. Binding of the complex to this structure may then induce a conformational change and release the E3 ligase activity of FANCL resulting in monoubiquitylation of FANCD2 and activation of the FA pathway. In the near future, X-ray crystallography or single-particle electron microscopy techniques may give further insight into the function of the FA biochemical pathway and how this pathway protects the genome from DNA damage.

Evolution of the FANC core complex proteins: a novel mode of persistent rapid evolution? – In the previous chapters using a bioinformatic approach to assess the molecular function of the FANC gene products [Chapters 1–5], it was observed that the FANC proteins are considerably less conserved than average proteins in vertebrate species (Table 1). With fewer than 70% sequence identity between human and mouse, FANC proteins are amongst the 10% fastest evolving proteins [1, 2]. However, normal expectations based on evolutionary biology predict that a protein with a critical function should have a highly conserved sequence in order to maintain that function. Given the strong disadvantageous phenotype in humans, mice, and zebrafish when one of the *FANC* genes is absent or mutated [3–5], the extraordinarily high divergence between human, mouse, chicken and zebrafish FANC sequences is therefore unexpected.

Estimating the rate of non-synonymous substitutions (*Ka*) versus that of synonymous substitutions (*Ks*) between coding sequences of closely related species using algorithms such as the K-Estimator program [6], may give insight to the selective forces that act on an evolving protein coding region. A protein coding region, such as a recently duplicated gene, that not (yet) supports a vital function ('relaxed constraint') will evolve neutrally, with a *Ka*/*Ks* ratio $\approx$ 1. Most genes, however, have already acquired an essential function and most amino acid substitutions will therefore be detrimental and these genes will evolve under negative selection (*Ka*/*Ks* << 1). Positive selection, as indicated by *Ka*/*Ks* > 1, occurs

when protein functionality is not optimal anymore due to a changing environment and indicates that a protein coding region is under adaptive evolution.

Table 1. Estimation of the percentage identity between the human FANC proteins and their mouse, chicken, zebrafish and fruit fly orthologs

Protein	Number of residues	% Identity human-mouse	% Identity human-chicken	% Identity human - zebrafish	% Identity human-fruit fly
FANCA	1455	66	~46	?	-
FANCB	859	~49	~42	~25	-
FANCC	558	67	49	~28	-
FANCD2	1471	75	56	53	23
FANCE	536	~61	~38	~30	-
FANCF	374	~49	?	31	-
FANCG	622	72	36	26	-
FANCL	375	79	70	59	22
Histone H4	103	100%	100%	100%	98%
Cytochrome C	105	91%	87%	84%	78%
Hemoglobin α	142	86%	70%	56%	-

Note that ~ indicates that the estimation was based on a partial sequence only. For reference, three examples are given of proteins with various rates of evolution [7].

The *Ka*/*Ks* ratio for all FANC proteins is about 0.35 (Table 2) Although disease genes have previously been reported to have elevated *Ka/Ks* ratios compared to nondisease genes [8, 9], the *Ka/Ks* ratios of the *FANC* genes are significantly higher than for 'average' disease genes. Moreover, the higher than average *Ka/Ks* ratios are in agreement with the observed low percentage identities between human and mouse FANC protein sequences. As the actual numbers of neutral and adaptive substitutions can not be measured, the high *Ka/Ks* ratio of 0.35 can be attributed to either a relatively weak purifying selection (relaxed constraint) or to adaptive evolution of the FANC genes.

Biochemical data support the hypothesis that the FANC gene products have been evolving under *both* a weak purifying selection as well as adaptive evolution. Since no phenotypic difference exists between FA patients from the various complementation groups, all *FANC* genes acting upstream of *FANCD2* are known to be genetically epistatic. This indicates that the FA core complex itself is the 'functional unit' rather than the individual subunits. It has now been widely recognized that protein complexes are important functional units in cellular biochemistry. It can be speculated that in a highly iterative scenario, the individual subunits rapidly change over time but the overall structure and function of the core complex is maintained (Figure 1). In the 'first round', at interacting surfaces several substitutions that are neutral or nearly-neutral will accumulate. However, after these substitutions have occurred the interacting surfaces may be 're-optimized' in one or more subunits in the second round, possibly during a brief period of positive selection. In the third round additional (nearly) neutral substitutions are gathered

which will then again become compensated in the fourth round. And so on. In this system it becomes evident that amino acid positions that are functionally constraint in the first round may become neutral after a few rounds of evolution, and vice versa. This phenomenon has previously been described in the covarion model [reviewed in refs 10 and 11].

Table 2. Estimation of substitution rates of the FANC genes

Gene	Species compared	Acc. Numbers	No. of bp. analyzed	Ka	Ks	Ka/Ks
FANCA	human-mouse	NM_000135-NM_016925	4269	0,22	0,61	0,36
	mouse-rat	NM_016925-XM_341713	4070	0,075	0,21	0,36
FANCB	human-mouse	NM_152633-BC083151	1317	0,35	0,44	0,80
	mouse-rat	BC083151-BI302694	402	0,054	0,14	0,39
FANCC	human-mouse	NM_000136-NM_007985	1542	0,21	0,47	0,46
	mouse-rat	NM_007985-NM_012557	1659	0,045	0,12	0,37
FANCD2	human-mouse	NM_033084-XM_132796	4188	0,14	0,47	0,30
	mouse-rat	XM-132796-XM_232273	4101	0,040	0,17	0,23
FANCE	human-mouse	NM_021922-XM_283409	639	0,20	0,55	0,35
	mouse-rat	XM_283409-XM_228032	333	0,059	0,19	0,31
FANCF	human-mouse	NM_022725-AC113312	384	0,32	1,34	0,24
	mouse-rat	AC113312-CB580630	447	0,075	0,22	0,34
FANCG	human-mouse	BC000032-AY049715	1803	0,17	0,39	0,42
	mouse-chinese hamster	AY049715-AC084885	1848	0,090	0,25	0,36
FANCL	human-mouse	NM_018062-NM_025923	1125	0,12	0,40	0,31
	mouse-rat	NM_025923-XM_223701	1095	0,045	0,24	0,19

Mouse Fanca, -c, and -g orthologs can act as substitutes for their human counterparts, despite their low levels of identity of 66–72% with human [12–14]. This indicates that a high fraction of neutral substitutions must be possible. On the other hand, the *Xenopus laevis* Fancf and zebrafish Fancg proteins (27% and 26% identity with human) are not able to substitute their human core complex components [15, E. Blom unpublished result] showing that *Xenopus* and zebrafish FANC protein 3D structures are significantly different from human. This altered structure must be compensated in one or more other FANC core complex subunits and brief moments of adaptation must have occurred in order to maintain the overall function of the complex during evolution. Furthermore, the conservation of the overall structure and function of the FANC core complex has been illustrated in zebrafish. Fancd2-deficient zebrafish can be corrected by injection of the *human* FANCD2 mRNA [5]. This indicates that the zebrafish core complex accurately recognizes its human FANCD2 ortholog and activates it accordingly to result in restoration of the FA pathway.

The scenario presented here appears to be similar as for previously reported compensated pathogenic deviations (CPDs) [16, 17]. It has been shown that mutations that are known to be pathogenic in one species occasionally become fixed in a phylogenetic closely related species. For such mutation to become the wild type sequence, second-site or compensatory mutations must have co-evolved.

Although it has been suggested that intramolecular compensation predominates for CPDs [17], it is likely that for the FANC core complex evolution intermolecular compensation must have contributed greatly.

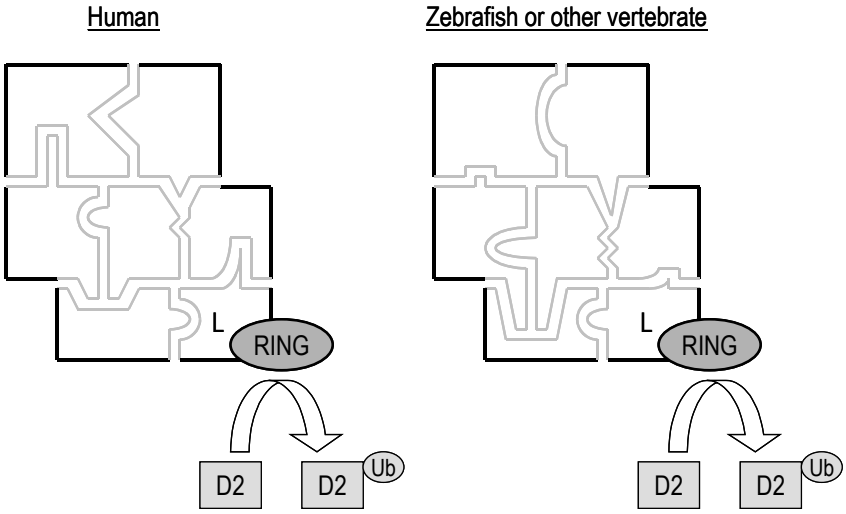


Fig. 1. Model of the human and zebrafish FANC core complexes, composed of the subunits FANCA, -B, -C, -E, -F, -G, and -L. The FANCL subunit contains the enzymatic RING domain that is required for activation of FANCD2 by monoubiquitylation. The black lines indicate that the overall shape and structure remains identical during evolution whereas the interacting surfaces between each subunit (grey lines) evolve rapidly and are highly different between human and zebrafish.

The above evolutionary scenario of the FANC complex may be important for several reasons. Firstly, it may well be of relevance for other protein complexes. Although most protein complexes contain several highly conserved domains, the protein-protein interaction regions may be subjected to the same mode of rapid evolution as shown for the FANC genes. Secondly, the model presented here may help to explain how 'orphan genes' arise during evolution. Orphan proteins are gene products that have no detectable similarity to other proteins in the database. Interestingly, FANCA, -B, -C, -E, and -F are currently all classified as orphans. In contrast to neutral (relaxed constraint) and adaptive evolution, which may only occur for a brief period during the history of an evolving protein [18], the scenario proposed here is intrinsically 'endless'. Assuming that the FANC genes all have arisen from duplication from an ancestral gene, the phenomenon of intrinsic continuity of change may explain why for the FANC genes the homology with their ancestral gene or paralog has been lost, in contrast to non-orphan genes which have only been subjected to change for relatively short times during their evolutionary history.

Acknowledgements
I would like to thank Rolf Hoekstra for his help and for giving me the opportunity to present and discuss the here presented evolution data at his lab at the Wageningen University in June 2005.

References

1 Makalowski W, Boguski MS (1998). Evolutionary parameters of the transcribed mammalinan genome: an analysis of 2,820 orthologous rodent and human sequences. *Proc Natl Acad Sci USA* **95:** 9407–9412.
2 Swanson WJ, Vacquier VD (2002). The rapid evolution of reproductive proteins. *Nature Rev Genet* **3:** 137–144.
3 Tischkowitz MD, Hodgson SV (2003). Fanconi anaemia. *J Med Genet* **40:** 1–10.
4 Wong JC, Buchwald M (2002). Disease model: Fanconi anemia. *Trends Mol. Med.* **8:** 139–142.
5 Liu TX *et al.* (2003). Knockdown of zebrafish Fancd2 causes developmental abnormalities via p53-dependent apoptosis. *Dev Cell.* **5:** 903–914.
6 Comeron JM (1999). K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* **15:** 763–764.
7 Alberts B, *et al.* (1994). Molecular biology of the cell. Third Edition. Garland Publishing New York & London. Chapter 6, p. 243–244.
8 Smith NGC, Eyre-Walker A (2003). Human disease genes: patterns and predictions. *Gene* **318:** 169–175.
9 Huang H *et al.* (2004). Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. *Genome Biol* **5:** R47.
10 Miyamoto MM, Fitch WM (1995). Testing the covarion hypothesis of molecular evolution. *Mol Biol Evol* **12:** 503–513.
11 Penny D, McComish BJ, Charleston MA, Hendy MD (2001). Mathematecal elegance with biochemical realism: the covarion model of molecular evolution. *J Mol Evol* **53:** 711–723.
12 Van de Vrugt HJ, *et al.* (2000). Cloning and characterization of murine Fanconi anemia group A gene: Fanca protein is expressed in lymphoid tissues, testis, and ovary. *Mamm Genome* **11:** 326–331.
13 Wevrick R, Clarke CA, Buchwald M (1993). Cloning and analysis of the murine Fanconi anemia group C cDNA. *Hum Mol Genet* **2:** 655–662.
14 Van de Vrugt HJ, *et al.* (2002). Characterization, expression and complex formation of the murine Fanconi anaemia gene product Fancg. *Genes Cells* **7:** 333–342.

15 Léveillé F, et al. (2004). The Fanconi anemia gene product FANCF is a flexible adaptor protein. *J Biol Chem* **279:** 39421–39430.
16 Kulathinal RJ, Bettencourt BR, Hartl DL (2004). Compensated deleterious mutations in insect genomes. *Science* **306:** 1553–1554.
17 Kondrashov AS, Sunyaev S, Kondrashov FA (2002). Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci USA* **99:** 14878–14883.
18 Jordan IK, Wolf YI, Koonin EV (2004). Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol Biol* **4:** 22.

Nederlandse samenvatting

Het ontrafelen van de FA biochemische route

Onderzoek op het grensvlak van biochemie, moleculaire evolutiebiologie en bioinformatica

Fanconi anemie (afgekort FA) is een erfelijke ziekte die in 1927 voor het eerst beschreven werd door de Zwitserse kinderarts Guido Fanconi. De aandoening Fanconi anemie kenmerkt zich door een zeer ernstige vorm van bloedarmoede, een sterk verhoogd risico op het krijgen van kanker (voornamelijk acute myeloïde leukemie en plaveiselcel-carcinoom) en aangeboren afwijkingen zoals het ontbreken van de duim. Het falen van het beenmerg om voldoende bloedcellen aan te maken en het verhoogde kankerrisico maken dat FA patiënten gemiddeld niet ouder worden dan 25 jaar. Op dit moment is de belangrijkste vorm van therapie voor FA patiënten de beenmergtransplantatie, maar dit kan de ziekte helaas niet volledig genezen.

FA is genetisch zeer heterogeen, wat wil zeggen dat op dit moment aangenomen wordt dat tenminste 11 genen een rol spelen bij het ontstaan van de ziekte. Van deze 11 genen zijn er reeds 9 geïdentificeerd: *FANCA, B, C, D1, D2, E, F, G,* en *L*. Biochemische studies hebben aangetoond dat the FANC genproducten met elkaar samenwerken in één biochemische 'route'. In dit proces verzamelen zich de meeste FANC eiwitten in een eiwitcomplex (dat zijn FANCA, B, C, E, F, G en L) wat het FA kerncomplex genoemd wordt. De aanwezigheid van dit complex is vervolgens verantwoordelijk voor de activatie van het FANCD2 eiwit door monoubiquitylering van het aminozuur Lysine[561]. Deze activatie induceert volgende biochemische reacties zoals associatie van FANCD2 met het BRCA1 eiwit in 'nuclear repair foci'. Het FANCD1 eiwit, dat hetzelfde is als het borstkanker genproduct BRCA2, speelt ook een rol in de FA biochemische route maar dit eiwit komt waarschijnlijk pas in actie na de activering van FANCD2.

Omdat cellen van FA patiënten hypergevoelig zijn voor stoffen die een covalente binding veroorzaken tussen de twee ketens van de dubbele DNA helix ('DNA interstrand cross-link'), wordt algemeen aangenomen dat de FA biochemische route een belangrijke rol speelt bij de reparatie van dit type DNA schade. Opmerkelijk is dat de DNA interstrand cross-link één van de meest schadelijke vormen van DNA schade is en, indien niet op een juiste manier gerepareerd, kan leiden tot het afsterven van de betreffende cel of tot ontstaan van kanker. Op dit moment is echter nog onduidelijk op welke wijze de FANC eiwitten bijdragen aan het reparatieproces van deze DNA interstrand cross-links. Een belangrijke reden hiervoor is dat de FANC eiwitten zogenaamde 'wees'-eiwitten zijn. In tegenstelling

tot de meeste andere genproducten lijken de FANC eiwitten compleet 'nieuwe' eiwitten. Ze zijn slecht geconserveerd tijdens de evolutie en vertonen geen enkele overeenkomst met andere eiwitten in de database. Het is daarom onmogelijk om op voorhand door middel van het toepassen van bioinformatica te voorspellen welke biochemische functie de FANC eiwitten hebben. Om dit probleem op te lossen werden de FANC eiwitsequenties in andere gewervelde diersoorten zoals kip, klauwpad en zebravis onderzocht. Tijdens de evolutie identiek gebleven aminozuren zijn waarschijnlijk belangrijk voor de functie en kunnen mogelijk inzicht geven in de moleculaire functie van de FANC eiwitten.

In **hoofdstuk 2** wordt het principe van 'wees'-eiwitten en dat van geconserveerde eiwitdomeinen verder uitgelegd. Verder wordt in dit hoofdstuk besproken hoe het toepassen van bioinformatica kan leiden tot meer inzicht in de functie van eiwitten. Uit een eerste analyse bleek dat FANCG toch geen weeseiwit is maar deel uitmaakt van de familie van tetratricopeptide repeat (TPR) eiwitten. Dit werd verder onderzocht in **hoofdstuk 3**, waarin in totaal 7 TPR motieven in het FANCG eiwit worden geïdentificeerd. Het TPR motief is een belangrijke eiwit-eiwit interactie module, die vaak een belangrijke rol speelt in de assemblage van mulitsubunit eiwitcomplexen. Mutatie-analyse experimenten tonen aan dat de TPR motieven in FANCG essentieel zijn voor een goed functioneren van de FA biochemische route en bevestigen dat deze motieven inderdaad nodig zijn voor eiwit-eiwit interacties. In **hoofdstuk 4** werd het FANCF eiwit bestudeerd maar in de sequentie werd geen eiwitdomein gevonden. Een uitgebreide biochemische studie toont echter aan dat dit eiwit waarschijnlijk functioneert als 'adaptor'-molecuul in het FA kerncomplex. De overige FANC eiwitsequenties werden bestudeerd in **hoofdstuk 5**. Aanwijzingen werden gevonden dat FANCA en FANCC beide deel uitmaken van de familie van HEAT-repeat eiwitten. Op basis van dit gegeven is het aannemelijk dat FANCA en FANCC – net als het FANCG eiwit – een belangrijke rol spelen bij de assemblage van het FA kerncomplex. Verder werd bewijs gevonden dat FANCD2 waarschijnlijk dezelfde eigenschappen bezit als dat van coiled-coil eiwitten. In tegenstelling tot wat is voorgesteld in een publicatie van Meetei *et al.* (2003) heeft het FANCL eiwit geen PHD-finger domein maar in plaats daarvan een RING-domein. Door de aanwezigheid van dit RING-domein bezit FANCL een E3 ligase enzymatische activiteit en is het in staat om FANCD2 te monoubiquityleren.

De FA biochemische route werd bestudeerd in de fruitvlieg *Drosophila melanogaster*. Bijzonder is dat het genoom van de fruitvlieg slechts twee *FANC* genen bevat: het enzymatische gedeelte van het FA kerncomplex *dmFANCL* en het substraat van dit enzym, *dmFANCD2*. Dit betekent waarschijnlijk dat de fruitvlieg een eenvoudiger vorm van de FA biochemische route bezit. Om deze reden is dit

organisme interessant als model om de functie FANC eiwitten beter te kunnen begrijpen. De eerste resultaten over dit onderwerp worden besproken in **hoofdstuk 6**.

De resultaten in dit proefschrift tonen aan dat het toch mogelijk is om weeseiwitten, zoals FANCG, FANCA, FANCC en FANCD2, te classificeren in een reeds bekende familie van eiwitdomeinen. Dit betekent dat de meeste weeseiwitten waarschijnlijk ontstaan zijn tijdens de evolutie door duplicatie van een reeds bestaand gen. Een alternatief mechanisme waarbij willekeurig voorkomende open reading frames (ORFs) in het genoom gebruikt worden om nieuwe genen te laten ontstaan zal om deze reden minder belangrijk zijn voor FANC eiwitten. De ontwikkeling en toepassing van gevoeliger database zoekalgoritmen zal waarschijnlijk leiden tot een afname van het aantal weeseiwitsequenties in de database. Op dit moment zijn ongeveer 10 tot 30 procent van alle voorspelde open reading frames geclassificeerd als weeseiwitten.

Opmerkelijk is dat behalve voor FANCL geen enkel eiwit uit het FA kerncomplex een enzymatisch domein bezit. Dit betekent waarschijnlijk dat de totale ruimtelijke structuur van het complex samen met mogelijke conformatieveranderingen het belangrijkste kenmerk van het complex zal zijn. Mogelijk vertoont om deze reden het FA kerncomplex veel overeenkomst met het ATM eiwit. ATM is een zeer groot eiwit van 3056 aminozuren en is, op een klein PI3-kinase domein in de C-terminus na, volledig samengesteld uit HEAT-repeats. Single-particle electron microscopy experimenten die zijn uitgevoerd door Llorca *et al.* (2003) hebben aangetoond dat ATM is samengesteld uit een 'head' en 'arm' domein. Na binding van ATM aan DNA verandert de conformatie van het 'arm' domein en wikkelt deze zich om de DNA keten heen. Het FA kerncomplex functioneert mogelijk op een vergelijkbare wijze. De ruimtelijke structuur van het complex zou een bepaalde DNA structuur kunnen herkennen, mogelijk 'stalled replication forks' of andere tussenvormen van het DNA herstelproces. Binding van het FA kerncomplex aan deze DNA structuur induceert vervolgens een conformatie verandering waarna de E3 ligase activiteit van FANCL vrijkomt en het FANCD2 geactiveerd wordt door monoubiquitylering. In de nabije toekomst zal de toepassing van röntgen diffractie, NMR en single-particle electron microscopy technieken een bijdrage leveren aan het verder ophelderen van de functie van de FA biochemische route en op welke wijze deze route bijdraagt aan de bescherming van het genoom tegen DNA veranderingen.

Curriculum vitae, publicatielijst en dankwoord

Curriculum vitae

Eric Blom werd geboren op 10 februari 1978 te Woerden. Hij groeide op in het Twentse Hengelo en behaalde daar in 1996 zijn Atheneum diploma aan de Openbare Scholengemeenschap Bataafse Kamp. In datzelfde jaar begon hij zijn studie Biomedische Wetenschappen aan de Universiteit Leiden. Tijdens zijn studie liep hij stage bij de afdelingen Moleculaire Celbiologie en Maag-, Darm- en Leverziekten van het Leids Universitair Medisch Centrum (LUMC). Zijn afstudeerstage en scriptie werden eveneens in het LUMC uitgevoerd bij de afdeling Anatomie en Embryologie onder de enthousiaste begeleiding van Dr. B.P. Hierck. Na het behalen van zijn bul eind 2000 begon hij hetzelfde jaar nog met zijn promotieonderzoek naar de moleculaire functie van Fanconi anemie genen bij de afdeling Klinische Genetica en Antropogenetica aan het VU medisch centrum in Amsterdam. Hij werd daarbij begeleid door Prof.dr. H. Joenje, Dr. F. Arwert en Dr. H.J. van de Vrugt. De resultaten van dit onderzoek zijn in dit proefschrift beschreven. Eric Blom is sinds juni 2005 bij dezelfde afdeling werkzaam als post-doc onderzoeker.

Publicatielijst

Van de Vrugt HJ, Koomen M, Berns MA, De Vries Y, Rooimans MA, Van der Weel L, Blom E, De Groot J, Schepers RJ, Stone S, Hoatlin ME, Cheng NC, Joenje H, Arwert F (2002). Characterization, expression and complex formation of the murine Fanconi anaemia gene product Fancg. *Genes Cells* **7**: 333–342.

Blom E, Van de Vrugt HJ, De Winter JP, Arwert F, Joenje H (2002). Evolutionary clues to the molecular function of Fanconi anemia genes. *Acta Haematol* **108**: 231–236.

Kuyvenhoven JPh, Van Hoek B, Blom E, Van Duijn W, Hanemaaijer R, Verheijen JH, Lamers CBHW, Verspaget HW (2003). Assessment of the clinical significance of serum matrix metalloproteinases MMP-2 and MMP-9 in patients with various chronic liver diseases and hepatocellular carcinoma. *Thromb Haemost* **89**: 718–725.

Blom E, Van de Vrugt HJ, De Vries Y, De Winter JP, Arwert F, Joenje H (2004). Multiple TPR motifs characterize the Fanconi anemia FANCG protein. *DNA Repair* **3**: 77–84.

Léveillé F, Blom E, Medhurst AL, Bier P, Laghmani EH, Johnson M, Rooimans MA, Sobeck A, Waisfisz Q, Arwert F, Patel KJ, Hoatlin ME, Joenje H, De Winter JP (2004). Fanconi anemia gene product FANCF is a flexible adaptor protein. *J Biol Chem* **279**: 39421–29430.

Meyer S, Barber LM, White DJ , Will AM, Birch JM, Kohler JA, Ersfeld K, Blom E, Joenje H, Eden TOB, Taylor GM (2006). Spectrum and significance of variants and mutations in the Fanconi Anaemia Group G gene (FANCG) in children with sporadic acute myeloid leukaemia. *Br J Haematol, in press.*

Hussain S, Wilson JB, Blom E, Thompson LH, Sung P, Gordon SM, Kupfer GM, Joenje H, Mathew CG, Jones NJ (2006). Tetratricopeptide-motif-mediated interaction of FANCG with recombination proteins XRCC3 and BRCA2. *DNA repair, in press.*

Blom E, Arwert F, Hoekstra RF, Joenje H (2006). FANC core complex proteins are trapped in a novel mode of persistent rapid evolution. *Manuscript in preparation.*

Dankwoord

Het is af! Ik ben erg blij mijn proefschrift te hebben afgerond. Uiteraard was me dit niet gelukt zonder de hulp van velen. Bij deze wil ik iedereen bedanken die mij de afgelopen vijf jaar heeft geholpen.

Allereerst wil ik mijn huidige collega's en ex-collega's van de sectie Functionele Genoomanalyse van de afdeling Klinische Genetica en Antropogenetica van het VUmc bedanken. Anneke, Annette, Carola, Djoke, Dominique, Fei, France, Ilse, Laghmani, Mariska, Marta, Mireille, Najim, Patrick, Tini, Quinten en Yne bedankt voor de gezellige tijd bij de vakgroep, leuke labuitjes en koffiepauzes èn natuurlijk ook voor de vele tips en trucs die ik van jullie de afgelopen jaren heb gekregen. Jûrgen en Martin, meester celkwekers, van jullie leerde ik de fijne kneepjes van het 'ambacht' celkweken. Marieke, mede AIO, samen een werkkamer delen was zeker erg nuttig maar ook heel gezellig. Bedankt voor alle miniwerkbesprekingen en de handige tips bij het afronden van mijn proefschrift. Henri, post-doc onderzoeker, de eerste twee jaar van mijn onderzoek heb je me enorm geholpen bij het aanleren van de moleculair biologische technieken. Bedankt ook voor alle steun en voor je hulp bij het richting geven mijn onderzoek. Johan, mijn co-promotor, bedankt voor je support, de nuttige discussies en je aanstekelijke enthousiasme voor het wetenschappelijk onderzoek. Fré, mijn andere co-promotor, jij begeleidde de 'muizengroep' bestaande uit Henri, Mariska, Mireille en Yne. Toen ik daar bijkwam met mijn zebravis en Japanse rijstvis project werd dit ineens 'muizen en andere diersoorten groep'. Bedankt voor alle steun en wijze raad! Tenslotte Hans, mijn promotor, jouw support, ambitie en ook de vele gedachtewisselingen waren een onmisbare bijdrage bij het doen slagen van dit proefschrift. Bedankt!

Niet alleen op de VU maar ook vanuit mijn vriendenkring en familie kreeg ik de nodige aanmoedigen de afgelopen vijf jaar. Allereerst wil ik iedereen van onze 'BW groep' bedanken. We kennen elkaar sinds 1996 toen we met z'n allen met de studie BW in Leiden begonnen. In het begin waren we niet al te serieus (vooral niet tijdens de welbekende practica), maar dit werd steeds beter naarmate de studie vorderde. Ik ben heel blij dat we elkaar ook na onze studies nog zien. Diana, Jerry, Jessica, Joyce, Marissa en Simone bedankt! Alexandra, Bianca, Noor en Rob bedankt voor de vele gezellige etentjes, weekendjes weg, Sinterklaas en Oud en Nieuw vieringen en alle andere uitjes. Ik hoop dat we nog heel lang vrienden zullen blijven! Tenslotte, Nicolette, mijn beste vriendin en voor deze gelegenheid ook mijn paranimf. We kennen elkaar nu bijna 10 jaar. Ik word doctor en jij binnenkort dokter. Bedankt voor alle gezelligheid en vriendschap de afgelopen jaren! Tot slot, en zeker niet in de laatste plaats, wil ik mijn familie bedanken. Cor en Coby, Yvonne en Elmar, ook al

wonen jullie een beetje ver weg, daar in het Twentse land, jullie zijn toch al jaren een enorme steun voor me en zonder alle support was ik zeker niet zover gekomen.

Iedereen bedankt!

Eric